Acta Math., 211 (2013), 255–290
DOI: 10.1007/s11511-013-0104-z

Proof of the BMV conjecture

by

HERBERT R. STAHL

Beuth Hochschule
Berlin, Germany

1. Introduction

1.1. The conjecture

Let A and B be two $n\times n$ Hermitian matrices and let B be positive semidefinite. In [1] it has been conjectured that under these assumptions the function

$$f(t) := \operatorname{Tr} e^{A-tB}, \quad t \geqslant 0, \tag{1.1}$$

can be represented as the Laplace transform

$$f(t) = \int e^{-ts}\, d\mu_{A,B}(s) \tag{1.2}$$

of a positive measure $\mu_{A,B}$ on $\mathbb{R}_+ = [0,\infty)$. In the present article we prove this conjecture from 1975 and give a semi-explicit expression for the measure $\mu_{A,B}$ (cf. Theorems 1.3 and 1.6 below).

Over the years different approaches and techniques have been tested for proving the conjecture. Surveys are contained in [18] and [10]. Recent publications are typically concerned with techniques from non-commutative algebra and combinatorics ([11], [12], [9], [8], [10], [13], [14], [3], [6], [2]). This direction of research was opened by a reformulation of the problem in [15]. Although our approach will follow a different line of analysis, we nevertheless repeat the main assertions from [15] in the next subsection as points of reference for later discussions.

After this paper was accepted, the author sadly passed away in April 2013.

Research supported by the grant STA 299/13-1 der Deutschen Forschungsgemeinschaft (DFG).

1.2. Reformulations of the conjecture

Definition 1.1. A function $f \in C^\infty(\mathbb{R}_+)$ is called *completely monotonic* if

$$(-1)^m f^{(m)}(t) \geqslant 0 \quad \text{for all } m \in \mathbb{N} \text{ and } t \in \mathbb{R}_+.$$

By Bernstein's theorem about completely monotonic functions (cf. [4] or [21, Chapter IV]) this property is equivalent to the existence of the Laplace transform (1.2) with a positive measure on $\mathbb{R}_+$. In this way, Definition 1.1 gives a first reformulation of the BMV (Bessis–Moussa–Villani) conjecture.

In [15] two other reformulations have been proved. It has been shown that the conjecture is equivalent to each of the following two assertions:

(i) Let A and B be two positive semidefinite Hermitian matrices. For each $m \in \mathbb{N}$ the polynomial

$$t \longmapsto \operatorname{Tr}(A+tB)^m$$

has only non-negative coefficients.

(ii) Let A be a positive definite and B a positive semidefinite Hermitian matrix. For each $p>0$ the function

$$t \longmapsto \operatorname{Tr}(A+tB)^{-p}$$

is the Laplace transform of a positive measure on $\mathbb{R}_+$.

Especially, reformulation (i) has paved the way for extensive research activities with tools from non-commutative algebra; several of the papers have been mentioned earlier. The parameter m in assertion (i) introduces a new and discrete gradation of the problem. Presently, assertion (i) has been proved for $m \leqslant 13$ (cf. [8] and [13]). The BMV conjecture itself has remained unproven, even for the general case of matrices with a dimension as low as $n=3$. In his diploma thesis G. Grafendorfer [7] has investigated very carefully the case $n=3$ by a combination of numerical and analytical means, but no counterexample could be found.

In [15] one also finds a short review of the relevance of the BMV conjecture in mathematical physics, the area from which the problem arose originally.(1)

Among the earlier investigations of the conjecture, especially [17] has been very impressive and fascinating for the author. There, already in 1976, the conjecture was proved for a rather broad class of matrices, including the two groups of examples with explicit solutions that we will state next.

(1) Meanwhile, in a follow-up paper [16] to [15], the reformulations of the BMV conjecture have been extended, and the conjecture itself has been generalized by replacing the expression on the left-hand side of (1.1) by elementary symmetric polynomials of order $m \in \{1, ..., n\}$ of exponentials of the n eigenvalues of the expression $A-tB$. The expression in (1.1) with the trace operator then corresponds to the case $m=1$.

1.3. Two groups of examples with explicit solutions

1.3.1. Commuting matrices A and B

If the two matrices A and B commute, then they can be diagonalized simultaneously, and consequently the BMV conjecture becomes solvable rather easily; the measure $\mu_{A,B}$ in (1.2) is then given by

$$\mu_{A,B} = \sum_{j=1}^{n} e^{a_j} \delta_{b_j}, \tag{1.3}$$

with $a_1, ..., a_n$ and $b_1, ..., b_n$ being the eigenvalues of the two matrices A and B, respectively, and δ_x being the Dirac measure at the point x. Indeed, the trace of a matrix M is invariant under similarity transformations $M \mapsto TMT^{-1}$. Therefore, we can assume without loss of generality that A and B are given in diagonal form, and the measure (1.3) follows immediately.

1.3.2. Matrices of dimension $n=2$

We consider 2×2 Hermitian matrices A and B, with B assumed to be positive semidefinite. In order to keep notation simple, we assume B to be given in diagonal form $B=\operatorname{diag}(b_1, b_2)$ with $0 \leqslant b_1 \leqslant b_2$.

If $b_1 = b_2$, then, without loss of generality, also the matrix A can be assumed to be given in diagonal form, and consequently the case is covered by (1.3). Thus, we have to consider only the situation that

$$A = \begin{pmatrix} a_{11} & a_{12} \\ \bar{a}_{12} & a_{22} \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} b_1 & 0 \\ 0 & b_2 \end{pmatrix}, \quad 0 \leqslant b_1 < b_2 < \infty. \tag{1.4}$$

Proposition 1.2. *If the matrices A and B are given by (1.4), then the function $t \mapsto \operatorname{Tr}\exp(A - tB)$, $t \in \mathbb{R}_+$, in (1.1) can be represented as a Laplace transform (1.3) with the positive measure*

$$d\mu_{A,B}(t) = e^{a_{11}} d\delta_{b_1}(t) + e^{a_{22}} d\delta_{b_2}(t) + w_{A,B}(t) \chi_{(b_1,b_2)}(t)\, dt, \quad t \in \mathbb{R}_+, \tag{1.5}$$

where $\chi_{(b_1,b_2)}$ denotes the characteristic function of the interval (b_1, b_2), and the density function $w_{A,B}$ is given by

$$\begin{aligned} w_{A,B}(t) = {} & \frac{4}{(b_2-b_1)\pi} \exp\left(\frac{a_{11}(b_2-t)+a_{22}(t-b_1)}{b_2-b_1}\right) \\ & \times \int_0^{|a_{12}|} \cos\left(\frac{b_2+b_1-2t}{b_2-b_1} u\right) \sinh\left(\sqrt{|a_{12}|^2-u^2}\,\right) du. \end{aligned} \tag{1.6}$$

This density function is positive for all $b_1 < t < b_2$.

Proposition 1.2 will be proved in §7. In [17] an explicit solution has also been proved for dimension $n=2$; there the density function looks rather different from (1.6), and has the advantage that its positivity can be recognized immediately, while in our case of (1.6) a non-trivial proof of positivity is required (cf. §7.2).

1.4. The main result

We prove two theorems. In the first one, it is just stated that the BMV conjecture is true, while in the second one we give a semi-explicit representation for the positive measure $\mu_{A,B}$ in the Laplace transform (1.2). In many respects this second theorem is a generalization of Proposition 1.2.

THEOREM 1.3. *If A and B are two Hermitian matrices with B being positive semi-definite, then there exists a unique positive measure $\mu_{A,B}$ on $[0,\infty)$ such that (1.2) holds for $t\geqslant 0$. In other words, the BMV conjecture is true.*

For the formulation of the second theorem we need some preparations.

LEMMA 1.4. *Let A and B be the two matrices from Theorem 1.3. Then there exists a unitary matrix T_0 such that the transformed matrices $\widetilde{A}=(\tilde{a}_{ij}):=T_0^*AT_0$ and $\widetilde{B}:=T_0^*BT_0$ satisfy*

$$\widetilde{B}=\operatorname{diag}(\tilde{b}_1,...,\tilde{b}_n) \quad \text{with } 0\leqslant\tilde{b}_1\leqslant...\leqslant\tilde{b}_n, \tag{1.7}$$

and

$$\tilde{a}_{jk}=0 \quad \text{for all } j,k=1,...,n,\ j\neq k,\ \text{with } \tilde{b}_j=\tilde{b}_k. \tag{1.8}$$

Proof. The existence of a unitary matrix T_0 such that (1.7) holds is guaranteed by the assumption that B is Hermitian and positive semidefinite. If all $\tilde{b}_j$ are pairwise different, then requirement (1.8) is void. If however several $\tilde{b}_j$ are identical, then one can rotate the corresponding subspaces in such a way that in addition to (1.7) also (1.8) is satisfied. □

As the matrix $A-tB$ is Hermitian for $t\in\mathbb{R}_+$, there exists a unitary matrix $T_1=T_1(t)$ such that

$$T_1^*(A-tB)T_1=\operatorname{diag}(\lambda_1(t),...,\lambda_n(t)). \tag{1.9}$$

The n functions $\lambda_1,...,\lambda_n$ in (1.9) are restrictions to $\mathbb{R}_+$ of branches of the solution λ of the polynomial equation

$$g(\lambda,t):=\det(\lambda I-(A-tB))=0, \tag{1.10}$$

i.e., λ_j, $j=1,...,n$, is a branch of the solution λ if the pair $(\lambda,t)=(\lambda_j(t),t)$ satisfies (1.10) for each $t\in\mathbb{C}$. The solution λ is an algebraic function of degree n if the polynomial

$g(\lambda,t)$ is irreducible, and it consists of several algebraic functions otherwise. In the most extreme situation, the polynomial $g(\lambda,t)$ can be factorized into n linear factors, and this is exactly the case when the two matrices A and B commute, which has been discussed in §1.3.1.

In any case, the solution λ of (1.10) consists of one or several multivalued functions of t in $\mathbb{C}$, and the total number of different branches λ_j, $j=1,...,n$, is always exactly n. In the next lemma, properties of the functions λ_j, $j=1,...,n$, are assembled, which are relevant for the formulation of Theorem 1.6. The lemma will be proved in a slightly reformulated form as Lemma 3.3 in §3.

Lemma 1.5. *There exist n different branches λ_j, $j=1,...,n$, of the solution λ of* (1.10) *which are holomorphic in a punctured neighborhood of infinity. They can be numbered in such a way that we have*

$$\lambda_j(t)=\tilde{a}_{jj}-\tilde{b}_jt+O\left(\frac{1}{t}\right) \quad \text{as } t\to\infty, \quad j=1,...,n, \tag{1.11}$$

where the coefficients $\tilde{a}_{jj}$ and $\tilde{b}_j$, $j=1,...,n$, are elements of the matrices $\tilde{A}$ and $\widetilde{B}$ introduced in Lemma 1.4.

With Lemmas 1.4 and 1.5 we are ready to formulate the second theorem.

Theorem 1.6. *For the measure $\mu_{A,B}$ in* (1.3) *we have the representation*

$$d\mu_{A,B}(t)=\sum_{j=1}^{n} e^{\tilde{a}_{jj}}\,d\delta_{\tilde{b}_j}(t)+w_{A,B}(t)\,dt, \quad t\in\mathbb{R}_+, \tag{1.12}$$

with a density function $w_{A,B}$ that can be represented as

$$w_{A,B}(t)=\sum_{\tilde{b}_j<t}\frac{1}{2\pi i}\oint_{C_j} e^{\lambda_j(\zeta)+t\zeta}\,d\zeta \quad \text{for } t\in\mathbb{R}_+, \tag{1.13}$$

or equivalently as

$$w_{A,B}(t)=-\sum_{\tilde{b}_j>t}\frac{1}{2\pi i}\oint_{C_j} e^{\lambda_j(\zeta)+t\zeta}\,d\zeta \quad \text{for } t\in\mathbb{R}_+, \tag{1.14}$$

where each integration path C_j is a positively oriented, rectifiable Jordan curve in $\mathbb{C}$ with the property that the corresponding function λ_j is analytic on and outside of C_j. The values $\tilde{a}_{jj}$ and $\tilde{b}_j$, $j=1,...,n$, have been introduced in Lemma 1.4, and the functions λ_j, $j=1,...,n$, in Lemma 1.5.

The measure $\mu_{A,B}$ is positive, its support satisfies

$$\operatorname{supp}(\mu_{A,B}) \subseteq [\tilde{b}_1, \tilde{b}_n], \tag{1.15}$$

and the density function $w_{A,B}$ is a restriction of an entire function in each interval of $[\tilde{b}_1, \tilde{b}_n] \setminus \{\tilde{b}_1, ..., \tilde{b}_n\}$. In general the density function $w_{A,B}$ is not positive on each of the intervals in $[\tilde{b}_1, \tilde{b}_n] \setminus \{\tilde{b}_1, ..., \tilde{b}_n\}$, but the irreducibility of the polynomial $g(\lambda, t)$ in (1.10) is sufficient for $w_{A,B}(t) > 0$ for all $t \in [\tilde{b}_1, \tilde{b}_n] \setminus \{\tilde{b}_1, ..., \tilde{b}_n\}$.

Obviously, the non-negativity of the density function $w_{A,B}$ is, prima vista, not evident from representation (1.13) or (1.14); its proof will be the topic of §5.

The semi-explicit representation of the measure $\mu_{A,B}$ in Theorem 1.6 is of key importance for our strategy for a proof of the BMV conjecture, but it possesses also independent value. In any case, it already conveys some ideas about the nature of the solution.

1.5. Outline of the paper

Theorem 1.3 is practically a corollary of Theorem 1.6, and the proof of Theorem 1.6 is given in §§2–6.

We start in §2 with two technical assumptions, which simplify the notation, but do not restrict the generality of the treatment. After that, in §3 we compile and prove results concerning the solution λ of (1.10) and the associated complex manifold $\mathcal{R}_\lambda$, which is the natural domain of definition for λ.

In §4 all assertions in Theorem 1.6 are proved, except for the positivity of the measure $\mu_{A,B}$.

The proof of positivity of $\mu_{A,B}$ follows then in §5, and everything concerning the proofs of Theorems 1.3 and 1.6 is summed up in §6.

The proof of Proposition 1.2 follows in §7.

2. Technical assumptions

Assumption 1. Throughout §§3–6 we assume the matrices A and B to be given in the form (1.7) and (1.8) of Lemma 1.4, i.e., we have

$$B = \operatorname{diag}(b_1, ..., b_n) \quad \text{with } 0 \leqslant b_1 \leqslant ... \leqslant b_n < \infty, \tag{2.1}$$

and

$$a_{jk} = 0 \quad \text{for all } j, k = 1, ..., n, \ j \neq k, \text{ with } b_j = b_k. \tag{2.2}$$

Assumption 2. Further, we assume that

$$0 < b_1 \leqslant \ldots \leqslant b_n, \tag{2.3}$$

i.e., the matrix B is assumed to be positive definite.

Assumption 1 has the advantage that in the sequel we can write a_{jk} and b_j instead of $\tilde{a}_{jk}$ and $\tilde{b}_j$.

LEMMA 2.1. *Assumptions 1 and 2 do not restrict the generality of the proof of Theorems 1.3 and 1.6.*

Proof. In Lemma 1.4 it has been shown that there exists a similarity transformation $M \mapsto T_0^* M T_0$, with a unitary matrix T_0, such that any admissible pair of matrices A and B is transformed into matrices $\tilde{A}$ and $\widetilde{B}$ that have the special forms of (2.1) and (2.2). Since the trace of a matrix is invariant under such similarity transformations, we have

$$f(t) = \operatorname{Tr} e^{A-tB} = \operatorname{Tr} T_0^* e^{A-tB} T_0 = \operatorname{Tr} e^{T_0^* A T_0 - t T_0^* B T_0}$$

for all $t \in \mathbb{R}_+$, which shows that the function f in (1.1) remains invariant, and consequently the generality of the proofs of Theorems 1.3 and 1.6 is not restricted by Assumption 1.

If (2.3) is not satisfied, then the matrix $\widetilde{B} := B + \varepsilon I = \operatorname{diag}(\tilde{b}_1, \ldots, \tilde{b}_n)$ with $\varepsilon > 0$ satisfies Assumption 2. We have $\tilde{b}_j = b_j + \varepsilon$, $j = 1, \ldots, n$, and it follows from (1.1) that

$$\tilde{f}(t) := \operatorname{Tr} e^{A - t\widetilde{B}} = e^{-\varepsilon t} \operatorname{Tr} e^{A-tB} = e^{-\varepsilon t} f(t) \quad \text{for } t \geqslant 0. \tag{2.4}$$

From (2.4) and the translation property of Laplace transforms, we deduce that the measure $\mu_{A,B}$ in (1.2) for the function f is the image of the measure $\mu_{A,\widetilde{B}}$ for the function $\tilde{f}$ under the translation $t \mapsto t - \varepsilon$. Consequently, the proofs of Theorems 1.3 and 1.6 for the matrices A and $\widetilde{B}$ carries over to the situation with the original matrices A and B. □

3. Preparatory results

In the present section we compile some results and definitions that are concerned with the solution λ of the polynomial equation (1.10), and in addition we introduce a complex manifold $\mathcal{R}_\lambda$, which is the natural domain of definition of λ.

3.1. The branch functions $\boldsymbol{\lambda_1, \ldots, \lambda_n}$

The solution λ of the polynomial equation (1.10) is a multivalued function with n branches λ_j, $j = 1, \ldots, n$, defined in $\overline{\mathbb{C}}$. Each pair $(\lambda, t) = (\lambda_j(t), t)$ with $t \in \overline{\mathbb{C}}$, $j = 1, \ldots, n$, satisfies the equation

$$0 = g(\lambda, t) := \det(\lambda I - (A - tB)) = g_{(1)}(\lambda, t) \ldots g_{(m)}(\lambda, t), \tag{3.1}$$

which is identical to (1.10), with the difference that we now have added the polynomials $g_{(l)}(\lambda,t)\in\mathbb{C}[\lambda,t]$, $l=1,...,m$, which are assumed to be irreducible. If the polynomial $g(\lambda,t)$ itself is irreducible, then we have $m=1$, $g(\lambda,t)=g_{(1)}(\lambda,t)$, and λ is an algebraic function of order n. Otherwise, in case $m>1$, λ consists of m algebraic functions $\lambda_{(l)}$, $l=1,...,m$, which are defined by the m polynomial equations

$$g_{(l)}(\lambda_{(l)},t)=0, \quad l=1,...,m. \tag{3.2}$$

Hence, λ consists either of a single algebraic function or of several such functions, depending on whether $g(\lambda,t)$ is irreducible or not. In any case, the total number of branches λ_j is always exactly n.

Obviously, for each $t\in\mathbb{C}$, the numbers $\lambda_1(t),...,\lambda_n(t)$ are eigenvalues of the matrix $A-tB$, as has already been stated in (1.9). Since $A-tB$ is a Hermitian matrix for $t\in\mathbb{R}$, the restriction of each branch λ_j, $j=1,...,n$, to $\mathbb{R}$ is a real function.

From (3.1) and the Leibniz formula for determinants, we deduce that

$$g(\lambda,t)=\sum_{j=0}^{n} p_j(t)\lambda^j \tag{3.3}$$

with $p_j\in\mathbb{C}[t]$, $\deg p_j\leqslant n-j$ for $j=0,...,n$, $p_n\equiv 1$, and $p_{n-1}(t)=t\operatorname{Tr}B-\operatorname{Tr}A$. If $m>1$, then we assume that the polynomials $g_{(l)}$ are normalized by

$$g_{(l)}(\lambda,t)=\lambda^{n_l}+\text{lower terms in }\lambda, \quad l=1,...,m, \tag{3.4}$$

and we have $n_1+...+n_m=n$. In situations where we have to deal with individual algebraic functions $\lambda_{(l)}$, $l=1,...,m$, which will, however, not often be the case, we denote the elements of a complete set of branches of the algebraic function $\lambda_{(l)}$, $l=1,...,m$, by $\lambda_{l,k}$, $k=1,...,n_l$. There exists an obvious one-to-one correspondence

$$j:\{(l,k):k=1,...,n_l, l=1,...,m\}\longrightarrow\{1,...,n\}$$

such that the set of functions $\{\lambda_{l,k}:k=1,...,n_l,l=1,...,m\}$ bijectively corresponds to the set $\{\lambda_j:j=1,...,n\}$.

It is in the nature of branches of a multivalued function that their domains of definition possess a great degree of arbitrariness. Assumptions for limiting this freedom will be addressed in Definition 3.7 in the next subsection.

Since the solution λ of (3.1) consists either of a single or of several algebraic functions, it is obvious that λ possesses only finitely many branch points over $\overline{\mathbb{C}}$.

LEMMA 3.1. *All branches λ_j, $j=1,...,n$, of the solution λ of (3.1) can be chosen such that they are of real type, i.e. any function λ_j which is analytic in a domain $D_0\subset\mathbb{C}$, is also analytic in the domain $D_0\cup\{z:\bar{z}\in D_0\}$, and we have $\lambda_j(\bar{t})=\overline{\lambda_j(t)}$ for all $t\in D_0$.*

Proof. The relation $\lambda_j(\bar{t})=\overline{\lambda_j(t)}$ follows from the identity

$$\overline{g(\lambda,t)}=\det(\bar{\lambda}I-(\bar{A}-\bar{t}B))=\det(\bar{\lambda}I-(\bar{A}^t-\bar{t}B))=g(\bar{\lambda},\bar{t}),$$

which is a consequence of $\bar{A}^t=A^*=A$ and of B being diagonal. Since the restriction of λ_j to $\mathbb{R}$ is real, $\overline{\lambda_j(\bar{t})}$ is an analytic continuation of λ_j across $\mathbb{R}$. □

LEMMA 3.2. *The solution λ of (3.1) has no branch points over $\mathbb{R}$.*

Proof. The lemma is a consequence of the fact that the functions $\lambda_j, j=1,...,n$, are of real type. We give an indirect proof, and assume that $x_0\in\mathbb{R}$ is a branch point of order $k\geqslant 1$ of a branch λ_j, $j\in\{1,...,n\}$, which we may assume to be analytic in a slit neighborhood $V\setminus(i\mathbb{R}_-+x_0)$ of x_0. Using a local coordinate at x_0 leads to the function $g(u):=\lambda_j(x_0+u^{k+1})$, which is analytic in a neighborhood of $u=0$. Obviously, the function g is also of real type. Let $l_0\in\mathbb{N}$ be the smallest index in the development $g(u)=\sum_l c_l u^l$ such that $c_{l_0}\neq 0$ and $l_0\not\equiv 0 \pmod{k+1}$, which means that there exists $0<l_1\leqslant k$ with $l_0=m(k+1)+l_1$, $m\in\mathbb{N}$. Like $\lambda_j(z)=g((z-x_0)^{1/(k+1)})$, so also the modified function

$$\tilde{\lambda}_j(z):=\left[g((z-x_0)^{1/(k+1)})-\sum_{l=0}^{m}c_{l(k+1)}(z-x_0)^l\right](z-x_0)^{-m}$$

has a branch point of order k at x_0, and is of real type. We have

$$\tilde{\lambda}_j(z)=c_{l_0}(z-x_0)^{l_1/(k+1)}+\mathrm{O}((z-x_0)^{(l_1+1)/(k+1)})\quad\text{as } z\to x_0,$$

and consequently for $r>0$ sufficiently small we have

$$\left|\arg\tilde{\lambda}_j(x_0+re^{it})-\arg c_{l_0}-\frac{l_1}{k+1}t\right|\leqslant\frac{\pi}{4(k+1)}\quad\text{for all } 0\leqslant t\leqslant\pi,$$

which implies that

$$0<\frac{l_1-\frac{1}{2}}{k+1}\pi\leqslant|\arg\tilde{\lambda}_j(x_0+r)-\arg\tilde{\lambda}_j(x_0-r)|\leqslant\frac{l_1+\frac{1}{2}}{k+1}\pi<\pi. \tag{3.5}$$

As the function $\tilde{\lambda}_j$ is of real type, we have $\arg\tilde{\lambda}_j(x_0+r)\equiv 0 \pmod{\pi}$ and $\arg\tilde{\lambda}_j(x_0-r)\equiv 0 \pmod{\pi}$, which contradicts (3.5). □

Lemma 3.2 is covered by a classical theorem by F. Rellich (cf. [19, Theorem XII.3]).

Next, we investigate the behavior of the functions λ_j, $j=1,...,n$, in the neighborhood of infinity.

Lemma 3.3. *Let λ_j, $j=1,...,n$, denote n different branches of the solution λ of* (3.1). *This system of branches can be chosen in such a way that there exists a simply connected domain $U_\lambda \subset \overline{\mathbb{C}}$ with $\infty \in U_\lambda$ such that the following assertions hold*:

(i) *Each function λ_j, $j=1,...,n$, is defined throughout U_λ, and none of them has a branch point in U_λ.*

(ii) *The n functions λ_j, $j=1,...,n$, can be enumerated in such a way that at infinity we have*

$$\lambda_j(t) = a_{jj} - b_j t + O\left(\frac{1}{t}\right) \quad \text{as } t \to \infty, \tag{3.6}$$

with a_{jj} and b_j, $j=1,...,n$, being the diagonal elements of the matrices A and B, respectively, of (2.1) *and* (2.2) *in Assumption* 1.

Remark 3.4. Assumption 1 from §2 is decisive for the concrete form of (3.6), and (3.6) is decisive for the verification of the representation of the measure $\mu_{A,B}$ in Theorem 1.6, which will follow in §4.2 below. Notice that the similarity transformation $(A,B) \mapsto (\tilde{A}, \widetilde{B})$ from Lemma 1.4 in general changes the diagonal elements a_{jj}, $j=1,...,n$, of the matrix A, while it leaves the polynomial equation (3.1) and also the branches λ_j, $j=1,...,n$, invariant. For an illustration of the changes of the a_{jj}, $j=1,...,n$, one may consult (7.4), where the simple case of 2×2 matrices has been analysed.

Remark 3.5. With Assumption 1 from §2 it is obvious that Lemma 1.5 in §1.4 is a reformulation of Lemma 3.3.

Proof of Lemma 3.3. We first prove that the solution λ of (3.1) has no branch point over infinity, which then leads to a proof of assertion (i). The proof of assertion (ii) is more involved.

(i) As in the proof of Lemma 3.2 we prove the absence of a branch point at infinity indirectly, and assume that some function λ_j, $j \in \{1,...,n\}$, has a branch point of order $k \geqslant 1$ at infinity. The function λ_j is of real type, and as a branch of an algebraic function it has at most polynomial growth as $t \to \infty$. Hence, there exists $m_0 \in \mathbb{N}$ such that the function

$$\lambda_0(z) := z^{m_0} \lambda_j\left(\frac{1}{z}\right)$$

is bounded in a neighborhood of $x_0=0$. The function λ_0 is again of real type, and it has a branch point of order $k \geqslant 1$ at $x_0=0$.

After these preparations we can copy the reasoning in the proof of Lemma 3.2 line by line in order to show that our assumption leads to a contradiction.

From equation (3.1) together with (3.3), we further deduce that all n functions λ_j, $j=1,...,n$, are finite in $\mathbb{C}$.

Since the solution λ of (3.1) possesses only finitely many branch points and none at infinity, the branches $\lambda_1, ..., \lambda_n$ can be chosen in such a way that there exists a punctured neighborhood of infinity in which all n functions λ_j, $j=1, ..., n$, are defined and analytic, which concludes the proof of assertion (i).

At infinity the functions λ_j, $j=1, ..., n$, may have a pole. In the next part of the proof we shall see that this is indeed the case, and the pole is always simple.

(ii) The proof of (3.6) will be done in two steps. In the first one we determine a condition that has to be satisfied by the leading coefficient of the development of the function λ_j, $j=1, ..., n$, at infinity.

Let λ_0 denote one of the functions $\lambda_1, ..., \lambda_n$. From part (i) we know that there exists an open, simply connected neighborhood $U_0 \subset \overline{\mathbb{C}}$ of ∞ such that λ_0 is analytic in $U_0 \setminus \{\infty\}$ and meromorphic in U_0. Hence, λ_0 can be represented as

$$\lambda_0 = p + v \tag{3.7}$$

with a polynomial p and a function v analytic in U_0 with $v(\infty)=0$. We will show that the polynomial p is necessarily of the form

$$p(t) = c_0 - c_1 t \quad \text{with } c_1 \in \{b_1, ..., b_n\}. \tag{3.8}$$

The proof will be done indirectly, and we assume that

$$\deg p \neq 1 \quad \text{or} \quad p(t) = c_0 - c_1 t \ \text{ with } c_1 \notin \{b_1, ..., b_n\}. \tag{3.9}$$

From (3.9) and the assumption made with respect to v after (3.7), it follows that

$$|p(t) + b_j t - a_{jj} + v(t)| \to \infty \ \text{ as } t \to \infty \quad \text{for each } j = 1, ..., n. \tag{3.10}$$

From the definition of $g(\lambda, t)$ in (3.1) and the Leibniz formula for determinants we deduce that

$$\begin{aligned} g(\lambda_0(t), t) = \prod_{j=0}^{n} &(p(t) + b_j t - a_{jj} + v(t)) \\ &+ O\Big(\max_{j=1,...,n} |p(t) + b_j t - a_{jj} + v(t)|^{n-2} \Big) \quad \text{as } t \to \infty. \end{aligned} \tag{3.11}$$

Indeed, the product in (3.11) is built from the diagonal elements of the matrix

$$\lambda_0(t) I - (A - tB),$$

and any other term in the Leibniz formula contains at least two off-diagonal elements as factors, which leads to the error term in the second line of (3.11). From (3.9), (3.10), and Assumption 2 in §2 we deduce that

$$\lim_{t\to\infty}\frac{|p(t)+b_kt-a_{kk}+v(t)|}{\max_{j=1,\dots,n}|p(t)+b_jt-a_{jj}+v(t)|}>0\quad\text{for each } k=1,\dots,n,$$

which implies that

$$\max_{j=1,\dots,n}|p(t)+b_jt-a_{jj}+v(t)|^{2-n}\prod_{j=0}^{n}|p(t)+b_jt-a_{jj}+v(t)|\to\infty \tag{3.12}$$

as $t\to\infty$. From (3.11) together with (3.10) and (3.12) it then follows that $g(\lambda_0(t),t)\to\infty$ as $t\to\infty$. But this contradicts $g(\lambda_0(t),t)=0$ for $t\in U_0$, and the contradiction proves the assertion made in (3.8).

We now come to the second step of the proof of (ii). Because of (3.8) we can make the ansatz

$$\begin{aligned}\lambda_j&=p_j+v_j &&\text{for } j=1,\dots,n,\\ p_j(t)&=c_{0j}-c_{1j}t &&\text{with } c_{1j}\in\{b_1,\dots,b_n\},\end{aligned}\tag{3.13}$$

where v_j is analytic in a neighborhood U_0 of infinity, and $v_j(\infty)=0$. We shall show that the functions $\lambda_1,\dots,\lambda_n$ can be enumerated in such a way that we have

$$c_{1j}=b_j\ \text{ and }\ c_{0j}=a_{jj}\quad\text{for each } j=1,\dots,n,$$

which proves (3.6).

A transformation of the variables λ and t into w and u is introduced by

$$u:=\frac{1}{t}\quad\text{and}\quad w:=\frac{1}{\lambda+b_1t-a_{00}},\tag{3.14}$$

with

$$a_{00}:=\min(\{c_{11},\dots,c_{1n},b_1,\dots,b_n\})-2.\tag{3.15}$$

From (3.14) it follows that

$$\lambda=\frac{1}{w}-b_1t+a_{00}=\frac{1}{w}-\frac{b_1}{u}+a_{00}.\tag{3.16}$$

There exists an obvious one-to-one correspondence between the n functions λ_j, for $j=1,\dots,n$, and the n functions

$$w_j(u):=\frac{1}{\lambda_j(1/u)+b_1/u-a_{00}},\quad j=1,\dots,n.\tag{3.17}$$

The functions w_j, $j=1,...,n$, are meromorphic in a neighborhood $\widetilde{U}_0$ of the origin. From (3.13) and (3.17) we deduce that

$$w_j(0)=\begin{cases} 0 & \text{for } c_{1j}\neq b_1, \\ \dfrac{1}{c_{0j}-a_{00}}\leqslant\dfrac{1}{2} & \text{for } c_{1j}=b_1, \end{cases} \tag{3.18}$$

and therefore we can choose $\widetilde{U}_0$ so small that

$$0<|w_j(u)|\leqslant 1 \quad \text{for } u\in\widetilde{U}_0\setminus\{0\}, \tag{3.19}$$

which implies that all w_j, $j=1,...,n$, are analytic in $\widetilde{U}_0$.

By $V(u)$, $u\in\mathbb{C}\setminus\{0\}$, we denote the $n\times n$ diagonal matrix

$$V(u):=\operatorname{diag}(\underbrace{1,...,1}_{m_1},\underbrace{\sqrt{u},...,\sqrt{u}}_{n-m_1}), \tag{3.20}$$

where m_1 is the number of appearances of b_1 in the multiset $\{b_1,...,b_n\}=\{b_j:j=1,...,n\}$, and we define

$$\tilde{g}(w,u):=\det(V(u)^2+w(B-b_1I)-wV(u)(A-a_{00}I)V(u)). \tag{3.21}$$

We then deduce that

$$\begin{aligned}\tilde{g}(w,u)&=\det\Big(V(u)\Big(I+\frac{w}{u}(B-b_1I)-w(A-a_{00}I)\Big)V(u)\Big)\\ &=w^nu^{n-m_1}\det\Big(\frac{1}{w}I+\frac{1}{u}(B-b_1I)-(A-a_{00}I)\Big)\\ &=w^nu^{n-m_1}\det\Big(\Big(\frac{1}{w}-\frac{b_1}{u}+a_{00}\Big)I-\Big(A-\frac{1}{u}B\Big)\Big)\\ &=w^nu^{n-m_1}g\Big(\lambda,\frac{1}{u}\Big).\end{aligned} \tag{3.22}$$

Indeed, the first equality is obvious if we take into account that

$$B-b_1I=\operatorname{diag}(0,...,0,b_{m_1+1}-b_1,...,b_n-b_1)$$

with exactly m_1 zeros in its diagonal. The next three equations follow from elementary transformations.

Directly from (3.21), but also from (3.3) and (3.22) together with (3.16), we deduce that $\tilde{g}(w,u)$ is a polynomial in w and u, and is of order n in w.

From (3.21) together with the properties used in (3.22) and the Leibniz formula for determinants it follows that, as $u\to 0$,

$$\tilde{g}(w,u)=\prod_{j=1}^{m_1}(1-w(a_{jj}-a_{00}))\prod_{j=m_1+1}^{n}(u-w(b_j-b_1)-wu(a_{jj}-a_{00}))(1+O(u)).\tag{3.23}$$

Indeed, the product in (3.23) is formed by the diagonal elements of the matrix

$$M:=V(u)^2+w(B-b_1I)-wV(u)(A-a_{00}I)V(u),$$

and the error term $O(u)$ in the second line of (3.23) follows from the fact that each other term in the Leibniz formula includes at least two off-diagonal elements of the matrix M as factors. Each off-diagonal element of M contains the factor $\sqrt{u}$, or it is zero since from Assumption 1 in §2 it follows that for all elements m_{jk} of $M=(m_{jk})$ with $j,k=1,...,m_1$, $j\neq k$, we have $m_{jk}=0$.

With (3.23) we are prepared to describe the behavior of the functions $w_1,...,w_n$ near $u=0$, which then translates into a proof of the first part of (3.6).

For each $u\in\mathbb{C}$ the n values $w_1(u),...,w_n(u)$ are the zeros of the polynomial

$$\tilde{g}(w,u)\in\mathbb{C}[w].$$

From (3.23) we know that

$$\tilde{g}(w,u)\to w^{n-m_1}\prod_{j=1}^{m_1}(1-w(a_{jj}-a_{00}))\prod_{j=m_1+1}^{n}(b_j-b_1)\quad\text{as } u\to 0.$$

Therefore it follows by Rouché's theorem that with an appropriate enumeration of the functions w_j, $j=1,...,n$, we have

$$\lim_{u\to 0}w_j(u)=\begin{cases}\dfrac{1}{a_{jj}-a_{00}} & \text{for } j=1,...,m_1,\\ 0 & \text{for } j=m_1+1,...,n,\end{cases}\tag{3.24}$$

which is a concretization of (3.18). Since we know from (3.19) that all functions w_j, $j=1,...,n$, are analytic in a neighborhood $\widetilde{U}_0$ of the origin, it follows from (3.24) that

$$w_j(u)=\frac{1}{a_{jj}-a_{00}}+O(u)\ \text{ as } u\to 0\quad\text{for } j=1,...,m_1.\tag{3.25}$$

From the correspondence (3.17) between the functions w_j and λ_j, it then follows from (3.25) that, for $j=1,...,m_1$,

$$\lambda_j(t)=\frac{1}{w_j(1/t)}-b_1t+a_{00}=a_{jj}-a_{00}-b_1t+a_{00}+O\left(\frac{1}{t}\right)=a_{jj}-b_jt+O\left(\frac{1}{t}\right)\tag{3.26}$$

as $t \to \infty$. The last equation is a consequence of $b_j = b_1$ for $j = 1, ..., m_1$. With (3.26) we have proved relation (3.6) for $j = 1, ..., m_1$.

By the definition of m_1 and the ordering in (2.3) we have

$$b_{m_1+1} > b_{m_1} = ... = b_1.$$

Let now m_2 denote the number of appearances of the value b_{m_1+1} in the multiset

$$\{b_j : j = 1, ..., n\}.$$

In order to prove relation (3.26) for $j = m_1+1, ..., m_1+m_2$, we repeat the analysis from (3.14) until (3.26) with b_1 replaced by b_{m_1+1} and m_1 by m_2, which then leads to the verification of (3.26) for $j = m_1+1, ..., m_1+m_2$.

Repeating this cycle for each different value b_j in the multiset $\{b_j : j = 1, ..., n\}$ proves relation (3.26) for all $j = 1, ..., n$, which completes the proof of (3.6), and concludes the proof of assertion (ii).

We would like to add as a short remark that if all b_j, $j = 1, ..., n$, are pairwise different, then the analysis in these last cycles could be considerably shortened since in such a case one could proceed rather directly from (3.18) to the conclusion (3.26). □

3.2. The complex manifold $\mathcal{R}_\lambda$

If the polynomial $g(\lambda, t)$ in (3.1) is irreducible, then the solution λ of (3.1) is an algebraic function of order n, and its natural domain of definition is a compact Riemann surface with n sheets over $\overline{\mathbb{C}}$ (cf. [5, Theorem IV.11.4]). We denote this surface by $\mathcal{R}_\lambda$.

If, however, the polynomial $g(\lambda, t)$ is reducible, then we have seen in (3.1) and (3.2) that the solution λ of (3.1) consists of m algebraic functions $\lambda_{(l)}$, $l = 1, ..., m$. Each $\lambda_{(l)}$ has a compact Riemann surface $\mathcal{R}_{\lambda,l}$, $l = 1, ..., m$, as its natural domain of definition, and therefore we have the disjoint union

$$\mathcal{R}_\lambda := \mathcal{R}_{\lambda,1} \cup ... \cup \mathcal{R}_{\lambda,m} \tag{3.27}$$

as the natural domain of definition for the multivalued function λ. In each of the two cases, $\mathcal{R}_\lambda$ is a covering of $\overline{\mathbb{C}}$ with exactly n sheets, except that in the latter case $\mathcal{R}_\lambda$ is no longer connected. Within each $\mathcal{R}_{\lambda,l}$, $l = 1, ..., m$, the different sheets are separated by cuts in the plane. By $\pi_\lambda : \mathcal{R}_\lambda \to \overline{\mathbb{C}}$ we denote the canonical projection of $\mathcal{R}_\lambda$.

A collection of subsets $\{S_\lambda^{(j)} \subset \mathcal{R}_\lambda : j = 1, ..., n\}$ forms a system of sheets on $\mathcal{R}_\lambda$ if the following three requirements are satisfied:

(i) The restriction $\pi_\lambda|_{S_\lambda^{(j)}} : S_\lambda^{(j)} \to \overline{\mathbb{C}}$ of the canonical projection π_λ is a bijection for each $j = 1, ..., n$.

(ii) We have $\bigcup_{j=1}^{n} S_\lambda^{(j)} = \mathcal{R}_\lambda$.

(iii) The interior points of each sheet $S_\lambda^{(j)} \subset \mathcal{R}_\lambda$, $j=1,...,n$, form a domain. Different sheets are disjoint except for branch points. A branch point of order $k \geqslant 1$ belongs to exactly $k+1$ sheets.

Because of requirement (i) each sheet $S_\lambda^{(j)}$ can be identified with $\overline{\mathbb{C}}$, however, formally we consider it as a subset of $\mathcal{R}_\lambda$.

While the association of branch points and sheets is specified completely in requirement (iii), there remains freedom with respect to the other boundary points of the sheets. We assume that this association is done in a pragmatic way. It is only required that each boundary point belongs to exactly one sheet if it is not a branch point.

Requirement (i) justifies the notational convention that a point of $S_\lambda^{(j)}$ is denoted by $t^{(j)}$ if $\pi_\lambda(t^{(j)}) = t \in \overline{\mathbb{C}}$.

The requirements (i)–(iii) give considerable freedom for choosing a system of sheets on $\mathcal{R}_\lambda$. In order to get unambiguity up to boundary associations, we define a standard system of sheets by the following additional requirement.

(iv) The cuts, which separate different sheets $S_\lambda^{(j)}$ in $\mathcal{R}_\lambda$, lie over lines in $\mathbb{C}$ that are perpendicular to $\mathbb{R}$. Each cut is chosen in a minimal way. Hence, it begins and ends with a branch point.

Lemma 3.6. *There exists a system of sheets $S_\lambda^{(j)} \subset \mathcal{R}_\lambda$, $j=1,...,n$, that satisfies the requirements* (i)–(iv). *Such a system is essentially unique, i.e., unique up to the association of boundary points that are not branch points. The domain U_λ from Lemma 3.3 can be chosen in such a way that each sheet $S_\lambda^{(j)}$, $j=1,...,n$, of the standard system covers U_λ, i.e., we have*

$$\pi_\lambda(\mathrm{Int}(S_\lambda^{(j)})) \supset U_\lambda. \tag{3.28}$$

Proof. From part (i) of Lemma 3.3 it is evident that there exist n unramified subdomains in $\mathcal{R}_\lambda$ over the domain U_λ; they are given by the set $\pi_\lambda^{-1}(U_\lambda)$. We can choose $U_\lambda \subset \overline{\mathbb{C}}$ as a disc around ∞. Because of Lemmas 3.1 and 3.2 it is then always possible to start an analytic continuation of a given branch λ_j, $j=1,...,n$, at ∞ and continue along rays that are perpendicular to $\mathbb{R}$ until one hits a branch point or the real axis. The first case can happen only finitely many times. Each of these continuations then defines a sheet $S_\lambda^{(j)}$, and the whole system satisfies the requirements (i)–(iv), and also (3.28) is satisfied. □

Each system $\{S_\lambda^{(j)} \subset \mathcal{R}_\lambda : j=1,...,n\}$ of sheets corresponds to a complete system of branches λ_j, $j=1,...,n$, of the solution λ of (3.1) if we define the functions λ_j by

$$\lambda_j := \lambda \circ \pi_{t,j}^{-1}, \quad j=1,...,n, \tag{3.29}$$

with $\pi_{\lambda,j}^{-1}$ denoting the inverse of $\pi_\lambda|_{S_\lambda^{(j)}}$, which exists because of requirement (i). If we use the standard system of sheets, then the branches λ_j, $j=1,...,n$, are uniquely defined functions.

Definition 3.7. In the sequel we denote by λ_j, $j=1,...,n$, the n branches of the solution λ of equation (3.1) that are defined by (3.29) with the standard system of sheets $\{S_\lambda^{(j)}: j=1\dots n\}$.

The next lemma is an immediate consequence of the monodromy theorem.

LEMMA 3.8. *Let λ_j, $j=1,...,n$, be the functions from Definition 3.7. Then for any entire function g the function*

$$G(t)=\sum_{j=1}^{n} g(\lambda_j(t)), \quad t\in\mathbb{C},$$

is analytic and single-valued throughout $\mathbb{C}$.

With the functions λ_j, $j=1,...,n$, we get a very helpful representation of the function f from (1.1) and also of the determinant $\det(\zeta I-(A-tB))$.

LEMMA 3.9. *With the functions λ_j, $j=1,...,n$, from Definition 3.7, the function f from (1.1) can be represented as*

$$f(t)=\operatorname{Tr} e^{A-tB}=\sum_{j=1}^{n} e^{\lambda_j(t)} \quad \textit{for } t\in\mathbb{C}. \tag{3.30}$$

It follows from Lemma 3.8 that f is an entire function.

Proof. From (3.1) it follows that for any $t\in\mathbb{C}$ the n numbers $\lambda_1(t),...,\lambda_n(t)$ are the eigenvalues of the the matrix $A-tB$. Let $V_\lambda\subset\mathbb{C}$ be the set of all $t\in\mathbb{C}$ such that not all $\lambda_1(t),...,\lambda_n(t)$ are pairwise different. This set is finite. For every $t\in\mathbb{C}\setminus V_\lambda$ the n eigenvectors corresponding to $\lambda_1(t),...,\lambda_n(t)$ form an eigenbasis. The $n\times n$ matrix $T_0=T_0(t)$ with these vectors as columns satisfies

$$T_0^{-1}(A-tB)T_0=\operatorname{diag}(\lambda_1(t),...,\lambda_n(t)). \tag{3.31}$$

Since the trace of a square matrix is invariant under similarity transformations, (3.30) follows from (3.31) and (1.1) for $t\notin V_\lambda$, and by continuity for all $t\in\mathbb{C}$. □

LEMMA 3.10. *With the functions λ_j, $j=1,...,n$, from Definition 3.7 we have*

$$\prod_{j=1}^{n}(\zeta-\lambda_j(t))=\det(\zeta I-(A-tB)) \quad \textit{for } \zeta,t\in\mathbb{C}. \tag{3.32}$$

Proof. From (3.31) we deduce that

$$T_0^{-1}(\zeta I-(A-tB))T_0=\operatorname{diag}(\zeta-\lambda_1(t),...,\zeta-\lambda_n(t))$$

for each $\zeta\in\mathbb{C}$ and $t\in\mathbb{C}\setminus V_\lambda$, which then proves (3.32). □

In the last lemma of this section we lift the complex conjugation from $\overline{\mathbb{C}}$ to $\mathcal{R}_\lambda$.

LEMMA 3.11. *There exists a unique anti-holomorphic mapping* $\varrho:\mathcal{R}_\lambda\to\mathcal{R}_\lambda$ *such that*

$$\pi_\lambda\circ\varrho(z)=\overline{\pi_\lambda(z)}\quad \text{for all } z\in\mathcal{R}_\lambda \tag{3.33}$$

and that $\varrho|_{\pi_\lambda^{-1}(\mathbb{R})}$ *is the identity.*

Proof. We start with the problem of existence. By requirement (i) of the standard system of sheets $\{S_\lambda^{(j)}:j=1...n\}$ on $\mathcal{R}_\lambda$, we can define ϱ on each $S_\lambda^{(j)}$, $j=1,...,n$, by a direct transfer of the complex conjugation from $\overline{\mathbb{C}}$ to $S_\lambda^{(j)}$. Note that each of the $\pi_\lambda(S_\lambda^{(j)})$, $j=1,...,n$, is invariant under complex conjugation because of requirement (iv) and since each λ_j is of real type. It is not difficult to see that this piecewise definition of ϱ is well defined throughout $\mathcal{R}_\lambda$, and possesses the required properties.

The uniqueness of ϱ is a consequence of the fact that $\varrho|_{\pi_\lambda^{-1}(\mathbb{R})}$ is the identity map. Indeed, let ϱ_1 and ϱ_2 be two maps with the required properties. Then $\varrho_1\circ\varrho_1$ and $\varrho_1\circ\varrho_2$ are both analytic maps from $\mathcal{R}_\lambda$ to $\mathcal{R}_\lambda$. On $\pi_\lambda^{-1}(\mathbb{R})$ both maps are the identity, and consequently $\varrho_1\circ\varrho_1$ and $\varrho_1\circ\varrho_2$ are both the identity map on $\mathcal{R}_\lambda$, which proves $\varrho_1=\varrho_2$. □

4. First part of the proof of Theorem 1.6

In this section we prove all assertions of Theorem 1.6 except for the positivity of the measure $\mu_{A,B}$, which will be the topic of the next section.

4.1. Equivalence of (1.13) and (1.14)

LEMMA 4.1. *For each* $t>0$ *we have*

$$\sum_{j=1}^{n}\frac{1}{2\pi i}\oint_{C_j}e^{\lambda_j(\zeta)+t\zeta}\,d\zeta=0 \tag{4.1}$$

with C_j *and* λ_j *as specified in Theorem 1.6.*

Proof. From Lemma 3.3 it is obvious that we can choose all C_j, $j=1,...,n$, to be identical to a single curve $C\subseteq\mathbb{C}$ such that all $\lambda_1,...,\lambda_n$ are analytic on and outside of C. We interchange summation and integration in (4.1), and deduce from Lemma 3.8 that $\sum_{j=1}^n e^{\lambda_j(\zeta)+t\zeta}=e^{t\zeta}\sum_{j=1}^n e^{\lambda_j(\zeta)}$ is an entire function, which proves (4.1). □

From (4.1) it follows immediately that the representations (1.13) and (1.14) in Theorem 1.6 for the density function $w_{A,B}$ are equivalent.

4.2. Proof of (1.12)–(1.14)

We use (1.12) and (1.13) in Theorem 1.6 as an ansatz for a measure $\mu_{A,B}$ and show by direct calculations that this measure satisfies (1.2).

From (1.13) it is evident that $w_{A,B}(t)=0$ for $0\leqslant t<b_1$; and since we know from the last subsection that (1.13) and (1.14) are equivalent representations, we further deduce from (1.14) that also $w_{A,B}(t)=0$ for $t>b_n$. From (1.12) and (1.13) we then get

$$\int e^{-ts}\,d\mu_{A,B}(s)=\sum_{j=1}^{n}e^{a_{jj}}e^{-tb_j}+\sum_{k=1}^{n-1}I_k(t), \tag{4.2}$$

with

$$I_k(t)=\int_{b_k}^{b_{k+1}}\sum_{j=1}^{k}\frac{1}{2\pi i}\oint_{C_j}e^{\lambda_j(\zeta)+s(\zeta-t)}\,d\zeta\,ds,\quad k=1,...,n-1. \tag{4.3}$$

As in the proof of Lemma 4.1, we assume again that all integration paths C_j, $j=1,...,n$, in (4.3) are identical with a single curve $C\subseteq\mathbb{C}$ such that all $\lambda_1,...,\lambda_n$ are analytic on and outside of C with a simple pole at infinity. Because of Lemma 3.2 we may assume that

$$\mathbb{R}_+\subset\operatorname{Ext}(C). \tag{4.4}$$

After these preparations we deduce from (4.3) that

$$\begin{aligned}\sum_{k=1}^{n-1}I_k(t)&=\sum_{k=1}^{n-1}\frac{1}{2\pi i}\oint_C e^{\lambda_k(\zeta)}\int_{b_k}^{b_n}e^{s(\zeta-t)}\,ds\,d\zeta\\&=\sum_{k=1}^{n-1}\frac{1}{2\pi i}\oint_C e^{\lambda_k(\zeta)}(e^{b_n(\zeta-t)}-e^{b_k(\zeta-t)})\frac{d\zeta}{\zeta-t}\\&=\sum_{k=1}^{n}\frac{-1}{2\pi i}\oint_C e^{\lambda_k(\zeta)}e^{b_k(\zeta-t)}\frac{d\zeta}{\zeta-t}\\&=\sum_{k=1}^{n}(e^{\lambda_k(t)}-e^{a_{kk}-tb_k}).\end{aligned} \tag{4.5}$$

Indeed, the first equality in (4.5) is a consequence of Fubini's theorem and (4.3), the second one follows from elementary integration, and the third one follows in the same way as the conclusion in the proof of Lemma 4.1. We give some more details, and deduce with the help of Lemma 3.8 that

$$\sum_{k=1}^{n}\frac{1}{2\pi i}\oint_C e^{\lambda_k(\zeta)}e^{b_n(\zeta-t)}\frac{d\zeta}{\zeta-t}=\frac{1}{2\pi i}\oint_C e^{b_n(\zeta-t)}\sum_{k=1}^{n}e^{\lambda_k(\zeta)}\frac{d\zeta}{\zeta-t}=0,$$

which then proves the third equality in (4.5). Notice that $t\in\operatorname{Ext}(C)$. For a verification of the last equality in (4.5) we define the functions r_k, $k=1,...,n$, by

$$\lambda_k(z)+b_k z=a_{kk}+r_k(z).$$

It then follows from (3.6) in Lemma 3.3 that $r_k(\infty)=0$ for $k=1,...,n$, and obviously each r_k is analytic on and outside of C. Since C is positively oriented, it follows from Cauchy's formula that

$$-\frac{1}{2\pi i}\oint_C e^{\lambda_k(\zeta)}e^{b_k(\zeta-t)}\frac{d\zeta}{\zeta-t}=-\frac{e^{-tb_k}}{2\pi i}\oint_C e^{\lambda_k(\zeta)+b_k\zeta}\frac{d\zeta}{\zeta-t}=-\frac{e^{a_{kk}-tb_k}}{2\pi i}\oint_C e^{r_k(\zeta)}\frac{d\zeta}{\zeta-t}$$
$$=e^{a_{kk}-tb_k}(e^{r_k(t)}-1)=e^{\lambda_k(t)}-e^{a_{kk}-tb_k}$$

for each $k=1,...,n$, which completes the verification of the last equality in (4.5).

By putting (4.2) and (4.5) together we arrive at (1.2), which proves that (1.12) and (1.13) is a representation of the measure $\mu_{A,B}$ that satisfies (1.2). From §4.1 it then follows that also (1.12) in combination with (1.14) defines the same measure $\mu_{A,B}$.

4.3. Proof of the inclusion (1.15)

Since before (4.2) we have verified that $w_{A,B}(t)=0$ for $0\leqslant t<b_1$ and for $t>b_n$, inclusion (1.15) in Theorem 1.6 follows from (1.12).

From (4.3) it is immediately obvious that the density function $w_{A,B}$ is the restriction of an entire function in each interval of the set $[b_1,b_n]\setminus\{b_1,...,b_n\}$.

4.4. Remark about the proof of (1.12)–(1.14)

In §4.2 the representation of the measure $\mu_{A,B}$ in Theorem 1.6 has been proved with the help of an ansatz. This strategy is very effective, but it gives no hints how one can systematically find such an ansatz. Actually, the expressions in (1.12) and (1.13) were only found after a lengthy asymptotic analysis of the function (1.1) with a subsequent application of the Post–Widder formulae for the inversion of Laplace transforms. Interested readers can find this systematic, but laborious, approach in [20].

5. The proof of positivity

For the completion of the proof of Theorem 1.6 it remains only to show that the measure $\mu_{A,B}$ is positive, which is done in this section. The essential problem is to show that the density function $w_{A,B}$ given by (1.13) or by (1.14) in Theorem 1.6 is non-negative in $[b_1,b_n]\setminus\{b_1,...,b_n\}$.

5.1. A preliminary assumption

In a first version of the proof of positivity we make the following additional assumption, which will afterwards, in §5.4, be shown to be superfluous.

Assumption 3. We assume that the polynomial $g(\lambda, t)$ in equation (3.1), which is identical to the polynomial in (1.10), is irreducible.

For the convenience of the reader we list the definitions from §3 that will be especially important in the next subsection. Some of them now have special properties because of Assumption 3.

(i) The solution λ of equation (3.1) is an algebraic function of degree n (cf. §3.1).

(ii) The covering manifold $\mathcal{R}_\lambda$ over $\overline{\mathbb{C}}$ from §3.2 is now a compact Riemann surface with n sheets over $\overline{\mathbb{C}}$. As before, by $\pi_\lambda: \mathcal{R}_\lambda \to \overline{\mathbb{C}}$ we denote its canonical projection.

(iii) The n functions λ_j, $j=1, ..., n$, from Definition 3.7 are n branches of the single algebraic function λ.

(iv) By C_j, $j=1, ..., n$, we denote n Jordan curves that are all identical with a single curve $C \subset \mathbb{C}$, and this curve is assumed to be smooth, positively oriented, and chosen in such a way that each function λ_j, $j=1, ..., n$, is analytic on and outside of C.

(v) The reflection function $\varrho: \mathcal{R}_\lambda \to \mathcal{R}_\lambda$ from Lemma 3.11 is the lifting of the complex conjugation from $\overline{\mathbb{C}}$ onto $\mathcal{R}_\lambda$, i.e., we have $\pi_\lambda(\varrho(\zeta)) = \overline{\pi_\lambda(\zeta)}$ for all $\zeta \in \mathcal{R}_\lambda$. By $\mathcal{R}_+ \subset \mathcal{R}_\lambda$ we denote the subsurface $\mathcal{R}_+ := \{z \in \mathcal{R}_\lambda : \operatorname{Im} \pi_\lambda(z) > 0\}$, and by $\mathcal{R}_- \subset \mathcal{R}_\lambda$ the corresponding subsurface defined over base points with a negative imaginary part; $\mathcal{R}_+$ and $\mathcal{R}_-$ are bordered Riemann surfaces over $\{z : \operatorname{Im} z > 0\}$ and $\{z : \operatorname{Im} z < 0\}$, respectively.

5.2. The main proposition

The proof of positivity under Assumption 3 is based on assertions that are formulated and proved in the next proposition.

Proposition 5.1. *Under Assumption* 3, *for any* $t \in (b_I, b_{I+1})$ *with* $I \in \{1, ..., n-1\}$ *there exists a chain* γ *of finitely many closed integration paths on the Riemann surface* $\mathcal{R}_\lambda$ *such that*

$$\operatorname{Im} e^{\lambda(\zeta) + t\pi_\lambda(\zeta)} = 0 \quad \text{for all } \zeta \in \gamma, \tag{5.1}$$

$$\frac{1}{2\pi i} \oint_\gamma e^{\lambda(\zeta) + t\pi_\lambda(\zeta)} \, d\zeta < 0, \tag{5.2}$$

$$\frac{1}{2\pi i} \oint_\gamma e^{\lambda(\zeta) + t\pi_\lambda(\zeta)} \, d\zeta = -\sum_{j=1}^{I} \frac{1}{2\pi i} \oint_{C_j} e^{\lambda_j(z) + tz} \, dz, \tag{5.3}$$

and as a consequence of (5.2) *and* (5.3) *we have*

$$\sum_{b_j<t} \frac{1}{2\pi i} \oint_{C_j} e^{\lambda_j(z)+tz}\, dz > 0. \tag{5.4}$$

The definition of the objects π_λ, λ, λ_j, C_j, $j=1,...,I$, *in* (5.1)–(5.4) *were listed in* (i)–(iv) *in the last subsection.*

The proof of Proposition 5.1 will be prepared by two lemmas and several technical definitions. Throughout this subsection the numbers $t\in(b_I, b_{I+1})$ and $I\in\{1,...,n-1\}$ are kept fixed, and Assumption 3 is effective.

We define

$$\begin{aligned} D_\pm &:= \{\zeta\in\mathcal{R}_\lambda : \pm\operatorname{Im}\pi_\lambda(\zeta)>0 \text{ and } \pm\operatorname{Im}(\lambda(\zeta)+t\pi_\lambda(\zeta))>0\}, \\ D &:= \operatorname{Int}(\overline{D_+\cup D_-}). \end{aligned} \tag{5.5}$$

The set $D\subset\mathcal{R}_\lambda$ is open, but not necessarily connected. Since the algebraic function λ is of real type, we have $\varrho(D_\pm)=D_\mp$ and $D_\pm\subset\mathcal{R}_\pm$ with the reflection function ϱ and Riemann surfaces $\mathcal{R}_+$ and $\mathcal{R}_-$ from (v) in the listing in the last subsection.

By $\mathrm{Cr}\subset\mathcal{R}_\lambda$ we denote the set of critical points of the function $\operatorname{Im}(\lambda+t\pi_\lambda)$, which are at the same time the critical points of $\operatorname{Re}(\lambda+t\pi_\lambda)$, and the zeros of the derivative $(\lambda+t\pi_\lambda)'$. Since $\mathcal{R}_\lambda$ is compact, it follows that Cr is finite.

LEMMA 5.2. (i) *The boundary* $\partial D\subset\mathcal{R}_\lambda$ *consists of a chain*

$$\gamma=\gamma_1+...+\gamma_K \tag{5.6}$$

of K *piecewise analytic Jordan curves* γ_k, $k=1,...,K$. *The orientation of each* γ_k, $k=1,...,K$, *is chosen in such a way that the domain* D *lies to its left. The curves* γ_k, $k=1,...,K$, *are not necessarily disjoint, however, intersections are possible only at critical points* $\zeta\in\mathrm{Cr}$.

(ii) *The choice of the Jordan curves* γ_k, $k=1,...,K$, *in* (5.6) *can be done in such a way that each of them is invariant under the reflection function* ϱ *except for its orientation, i.e., we have* $\varrho(\gamma_k)=-\gamma_k$ *for* $k=1,...,K$.

(iii) *Let* $2s_k$ *be the length of the Jordan curve* γ_k, $k=1,...,K$; *with a parametrization by arc length we then have* $\gamma_k\colon[0,2s_k]\to\partial D\subset\mathcal{R}_\lambda$. *The starting point* $\gamma_k(0)$ *can be chosen in such a way that*

$$\gamma_k((0,s_k))\subset\partial D_+\setminus\pi_\lambda^{-1}(\mathbb{R}) \quad \textit{and} \quad \gamma_k((s_k,2s_k))\subset\partial D_-\setminus\pi_\lambda^{-1}(\mathbb{R}). \tag{5.7}$$

(iv) *The function*

$$\operatorname{Re}(\lambda\circ\gamma_k+t(\pi_\lambda\circ\gamma_k))$$

is increasing on $(0,s_k)$, *decreasing on* $(s_k,2s_k)$, *and these monotonicities are strict at each* $\zeta\in\gamma_k\setminus(\mathrm{Cr}\cup\pi_\lambda^{-1}(\mathbb{R}))$.

Proof. The function $\mathrm{Im}(\lambda+t\pi_\lambda)$ is harmonic in $\mathcal{R}_\lambda\setminus\pi_\lambda^{-1}(\{\infty\})$. As a system of level lines of a harmonic function, ∂D consists of piecewise analytic arcs, and their orientations can be chosen in such a way that the domain D lies to the left of ∂D. Since $\partial D\setminus\mathrm{Cr}$ consists of analytic arcs, locally each $\zeta\in\partial D\setminus\mathrm{Cr}$ touches only two components of $\mathcal{R}_\lambda\setminus\partial D$, and locally it belongs only to one of the analytic Jordan subarcs of $\partial D\setminus\mathrm{Cr}$. Globally, for each $\zeta\in\partial D$ there exists at least one Jordan curve $\widetilde{\gamma}$ in ∂D with $\zeta\in\widetilde{\gamma}$, but this association is in general not unique, different choices may be possible, and the cuts that are candidates for such a choice bifurcate only at points in Cr. By a stepwise exhaustion it follows that ∂D is the union of Jordan curves, i.e., we have

$$\partial D=\gamma=\gamma_1+\gamma_2+\dots. \tag{5.8}$$

Different curves γ_k may intersect, but because of the implicit function theorem, intersections are possible only at points in Cr.

After these considerations it remains only to show in assertion (i) that the number of Jordan curves γ_k in (5.8) is finite; basically this follows from the compactness of $\mathcal{R}_\lambda$. If we assume that there exist infinitely many curves γ_k in (5.8), then there exists at least one cluster point $z^*\in\mathcal{R}_\lambda$ such that any neighborhood of z^* intersects infinitely many curves γ_k from (5.8). Obviously, $z^*\in\pi_\lambda^{-1}(\{\infty\})$ is impossible. Let $z\colon V\to\mathbb{D}$ be a local coordinate of z^* that maps a neighborhood V of z^* conformally onto the unit disc $\mathbb{D}$ with $z(z^*)=0$. The function

$$g:=\mathrm{Im}(\lambda+t\pi_\lambda)\circ z^{-1}$$

is harmonic in $\mathbb{D}$ and not identically constant. If g has a critical point of order m at the origin, then, by the local structure of level lines near a critical point, small neighborhoods of the origin can intersect only with at most m elements of the set $\{z(\gamma_k|_V):k=1,2,\dots\}$. If, on the other hand, g has no critical point at the origin, then it follows from the implicit function theorem that small neighborhoods of the origin can intersect with at most one element of the set $\{z(\gamma_k|_V):k=1,2,\dots\}$. Hence, the assumption that z^* is a cluster point of curves γ_k from (5.8) is impossible, and the finiteness of the sum in (5.8) is proved, which completes the proof of assertion (i).

For each Jordan curve γ_k, $k=1,\dots,K$, in (5.6) we deduce from (5.5) that

$$\frac{\partial}{\partial n}\,\mathrm{Im}(\lambda(\zeta)+t\pi_\lambda(\zeta))>0 \quad \text{for each } \zeta\in\gamma_k\cap(\mathcal{R}_+\setminus\mathrm{Cr}), \tag{5.9}$$

and since the orientation of $\partial D=\gamma$ has been chosen in such a way that D lies to the left of each γ_k, we further have

$$\frac{\partial}{\partial u}\,\mathrm{Re}(\lambda(\zeta)+t\pi_\lambda(\zeta))>0 \quad \text{for each } \zeta\in\gamma_k\cap(\mathcal{R}_+\setminus\mathrm{Cr}) \tag{5.10}$$

by the Cauchy–Riemann differential equations. In (5.9), $\partial/\partial n$ denotes the normal derivative on γ_k pointing into D, and in (5.10), $\partial/\partial u$ denotes the tangential derivative. In $\mathcal{R}_-$, we get the corresponding inequality

$$\frac{\partial}{\partial u}\operatorname{Re}(\lambda(\zeta)+t\pi_\lambda(\zeta))<0 \quad \text{for each } \zeta\in\gamma_k\cap(\mathcal{R}_-\setminus\mathrm{Cr}). \tag{5.11}$$

Since λ is a function of real type, we deduce with the help of the reflection function ϱ that

$$(\lambda\circ\varrho)(\zeta)+t(\pi_\lambda\circ\varrho)(\zeta)=\overline{\lambda(\zeta)+t\pi_\lambda(\zeta)} \quad \text{for } \zeta\in\mathcal{R}_\lambda,$$

and therefore also that

$$\varrho(\partial D)=\partial D. \tag{5.12}$$

As a first consequence of (5.10) and (5.11) we conclude that none of the Jordan curves γ_k in (5.6) can be contained completely in $\overline{\mathcal{R}}_+$ or $\overline{\mathcal{R}}_-$. Indeed, if we assume that some γ_k is contained in $\overline{\mathcal{R}}_+$, then it would follow from (5.10) that $\operatorname{Re}(\lambda+t\pi_\lambda)$ could not be continuous along the whole curve γ_k.

As each γ_k, $k=1,...,K$, in (5.6) intersects at the same time $\mathcal{R}_+$ and $\mathcal{R}_-$, it follows that all curves γ_k can be chosen from ∂D in the exhaustion process in the proof of assertion (i) in such a way that $\varrho(\gamma_k)=-\gamma_k$ for each $k=1,...,K$, which proves assertion (ii). We remark that a choice between different options for a selection of the γ_k, $k=1,...,K$, exists only if some points of the intersection $\gamma_k\cap\pi_\lambda^{-1}(\mathbb{R})$ belong to Cr.

From the fact that each γ_k in (5.6) is a Jordan curve, which is neither fully contained in $\overline{\mathcal{R}}_+$ nor in $\overline{\mathcal{R}}_-$ and that we have $\varrho(\gamma_k)=-\gamma_k$, we deduce that $\gamma_k\cap\pi_\lambda^{-1}(\mathbb{R})$ consists of exactly two points. By an appropriate choice of the starting point of the parametrization of γ_k in $\gamma_k\cap\pi_\lambda^{-1}(\mathbb{R})$ it follows that (5.7) is satisfied, which proves assertion (iii).

The monotonicity statements in assertion (iv) are immediate consequences of (5.10) and (5.11), which completes the proof of Lemma 5.2. □

Lemma 5.3. *We have*

$$\frac{1}{2\pi i}\oint_{\gamma_k} e^{\lambda(\zeta)+t\pi_\lambda(\zeta)}\,d\zeta<0 \quad \textit{for each } k=1,...,K. \tag{5.13}$$

Proof. We abbreviate the integrand in (5.13) by

$$g(\zeta):=e^{\lambda(\zeta)+t\pi_\lambda(\zeta)}, \quad \zeta\in\mathcal{R}_\lambda\setminus\pi_\lambda^{-1}(\{\infty\}),$$

and assume $k\in\{1,...,K\}$ in (5.13) to be fixed.

From assertion (i) in Lemma 5.2 we know that $\operatorname{Im} g(\zeta)=0$ for all $\zeta\in\gamma_k$, from assertion (iv) we further know that $\operatorname{Re} g(\zeta)=g(\zeta)$ is strictly increasing on $\gamma_k\cap(\mathcal{R}_+\setminus\mathrm{Cr})$, from (5.7) that $\gamma_k \cap \mathcal{R}_+$ is the subarc $\gamma_k|_{(0,s_k)}$, and from the proof of assertion (iv) it is evident that also the following slightly stronger statement

$$(g\circ\gamma_k)'(s) > 0 \quad \text{for } 0 < s < s_k \text{ and } \gamma_k(s) \notin \mathrm{Cr} \tag{5.14}$$

holds. It further follows from (5.7) that we have

$$\operatorname{Im} \pi_\lambda \circ \gamma_k(0) = \operatorname{Im} \pi_\lambda \circ \gamma_k(s_k) = 0 \quad \text{and} \quad \operatorname{Im} \pi_\lambda \circ \gamma_k(s) > 0 \text{ for } 0 < s < s_k. \tag{5.15}$$

Let the coordinates z, x and y and the differentials dz, dx and dy be defined by

$$\pi_\lambda(\zeta) = z = x + iy \in \mathbb{C}, \ \ \zeta \in \gamma_k, \quad \text{and} \quad dz = dx + i\,dy,$$

and let these coordinates and differentials be lifted from $\overline{\mathbb{C}}$ onto $\mathcal{R}_\lambda$, where we then have $\zeta = \xi + i\eta$ and $d\zeta = d\xi + i\,d\eta$. Taking into consideration that $\varrho(\gamma_k) = -\gamma_k$, $\varrho(d\zeta) = \overline{d\zeta}$, and $(g\circ\varrho)(\zeta) = \overline{g(\zeta)} = g(\zeta)$ for all $\zeta \in \gamma_k$, we conclude that

$$\begin{aligned} \frac{1}{2\pi i} \oint_{\gamma_k} g(\zeta)\,d\zeta &= \frac{1}{2\pi i} \int_{\gamma_k \cap D_+} g(\zeta)(d\xi + i\,d\eta) + \frac{1}{2\pi i} \int_{\gamma_k \cap D_-} g(\zeta)(d\xi + i\,d\eta) \\ &= \frac{1}{\pi} \int_{\gamma_k \cap D_+} g(\zeta)\,d\eta \\ &= \frac{1}{\pi} \int_0^{s_k} (g\circ\gamma_k)(s) \operatorname{Im}(\pi_\lambda \circ \gamma_k)'(s)\,ds \\ &= -\frac{1}{\pi} \int_0^{s_k} (g\circ\gamma_k)'(s) \operatorname{Im}(\pi_\lambda \circ \gamma_k)(s)\,ds < 0. \end{aligned} \tag{5.16}$$

Indeed, the first three equalities in (5.16) are a consequence of the specific symmetries and antisymmetries with respect to ϱ that have been listed just before (5.16). From the three equalities we consider the second one in more detail, and concentrate on the transformation of the second integral after the first equality. We have

$$\frac{1}{2\pi i} \int_{\gamma_k \cap D_-} g(\zeta)(d\xi + i\,d\eta) = -\frac{1}{2\pi i} \int_{\gamma_k \cap D_+} g(\zeta)(d\xi - i\,d\eta) = \frac{1}{2\pi i} \int_{\gamma_k \cap D_+} g(\zeta)(-d\xi + i\,d\eta),$$

which verifies the second equality. The last equality in (5.16) follows from integration by parts together with the equalities in (5.15). The inequality in (5.16) is then a consequence of (5.14) and the inequality in (5.15). □

Proof of Proposition 5.1. The chain γ of oriented Jordan curves (5.6) introduced in Lemma 5.2 is the candidate for the chain γ in Proposition 5.1. Equality (5.1) and inequality (5.2) have been verified by Lemmas 5.2 and 5.3, respectively. Identity (5.3) and its consequence (5.4) remain to be proved.

As integration paths C_j, $j=1,...,I$, on the right-hand side of (5.3) we take the common Jordan curve C from (iv) in the listing in the last subsection. The set $\pi_\lambda^{-1}(\overline{\mathrm{Ext}(C)})$ consists of n disjoint components if C is chosen sufficiently close to infinity; it then also follows that all branch points of λ are contained in $\mathcal{R}_\lambda \setminus \pi_\lambda^{-1}(\overline{\mathrm{Ext}(C)})$. Further, we have

$$\mathrm{Im}(\lambda_j(z)+tz)\begin{cases} >0 & \text{for all } z\in C \text{ with } \mathrm{Im}\, z>0, \\ <0 & \text{for all } z\in C \text{ with } \mathrm{Im}\, z<0, \end{cases} \qquad j=1,...,I, \tag{5.17}$$

and

$$\mathrm{Im}(\lambda_j(z)+tz)\begin{cases} <0 & \text{for all } z\in C \text{ with } \mathrm{Im}\, z>0, \\ >0 & \text{for all } z\in C \text{ with } \mathrm{Im}\, z<0, \end{cases} \qquad j=I+1,...,n. \tag{5.18}$$

A choice of C with these properties is possible because of (3.6) in Lemma 3.3 and the assumption that $b_1 \leqslant ... \leqslant b_I < t < b_{I+1} \leqslant ... \leqslant b_n$.

Next we define

$$D_0 := D \setminus \pi_\lambda^{-1}(\overline{\mathrm{Ext}(C)}) \subset \mathcal{R}_\lambda. \tag{5.19}$$

From (5.17), (5.18) and (5.5) it follows that exactly I of the n components $\widehat{C}_j \subset \mathcal{R}_\lambda$, $j=1,...,n$, of $\pi_\lambda^{-1}(\overline{\mathrm{Ext}(C)})$ are contained in D. Each $\widehat{C}_j$ lies in a different sheet $S_\lambda^{(j)}$, $j=1,...,n$, of the system of standard sheets introduced in Lemma 3.6. The enumeration of the sheets $S_\lambda^{(j)}$ corresponds to that of the functions λ_j as stated in (3.29). Let $\widetilde{C}_j \subset \mathcal{R}_\lambda$, $j=1,...,n$, denote the lifting of the oriented Jordan curve $C \subset \mathbb{C}$ onto $S_\lambda^{(j)} \subset \mathcal{R}_\lambda$. We then have $\pi_\lambda(\widetilde{C}_j)=C_j=C$ for $j=1,...,n$, and from (3.29) it follows that

$$\lambda(\zeta) = \lambda_j(\pi_\lambda(\zeta)) \text{ for } \zeta \in \widetilde{C}_j, \quad j=1,...,n. \tag{5.20}$$

Since $\widetilde{C}_j = \partial \widehat{C}_j$ for $j=1,...,n$, the open set D_0 lies to the left of each $\widetilde{C}_j$. Together with assertion (i) of Lemma 5.2, it follows from (5.19) that the chain

$$\gamma + \widetilde{C}_1 + ... + \widetilde{C}_I = \gamma_1 + ... + \gamma_K + \widetilde{C}_1 + ... + \widetilde{C}_I \subset \mathcal{R}_\lambda \tag{5.21}$$

forms the contour ∂D_0 with an orientation for which D_0 lies everywhere to its left. By Cauchy's theorem we have

$$\frac{1}{2\pi i} \oint_{\gamma+\widetilde{C}_1+...+\widetilde{C}_I} e^{\lambda(\zeta)+t\pi_\lambda(\zeta)}\, d\zeta = 0. \tag{5.22}$$

Identity (5.3) follows immediately from (5.22) and (5.20). Inequality (5.4) is a consequence of (5.2) and (5.3), since we have

$$\sum_{b_j<t} \frac{1}{2\pi i} \oint_{C_j} e^{\lambda_j(\zeta)+t\zeta}\, d\zeta = \sum_{j=1}^{I} \frac{1}{2\pi i} \oint_{C_j} e^{\lambda_j(\zeta)+t\zeta}\, d\zeta = -\frac{1}{2\pi i} \oint_\gamma e^{\lambda(\zeta)+t\pi_\lambda(\zeta)}\, d\zeta > 0. \tag{5.23}$$

□

5.3. A preliminary proof of positivity

With Proposition 5.1 we are prepared for the proof of positivity of the measure $\mu_{A,B}$ in Theorems 1.6 under Assumption 3, which then completes the proof of Theorem 1.6 under Assumption 3.

Proof of positivity under Assumption 3. From representation (1.12) in Theorem 1.6 it is obvious that the discrete part

$$d\mu_d = \sum_{j=1}^{n} e^{\tilde{a}_{jj}} \delta_{\tilde{b}_j} = \sum_{j=1}^{n} e^{a_{jj}} \delta_{b_j} \tag{5.24}$$

of the measure $\mu_{A,B}$ is positive. From (5.4) of Proposition 5.1 it follows that the density function $w_{A,B}$ in (1.13) of Theorem 1.6 is positive on

$$[\tilde{b}_1, \tilde{b}_n] \setminus \{\tilde{b}_1, ..., \tilde{b}_n\} = [b_1, b_n] \setminus \{b_1, ..., b_n\},$$

which proves the positivity of the measure $\mu_{A,B}$. Notice that the last identity holds because of Assumption 1 in §2. □

Under Assumption 3, relation (1.15) in Theorem 1.6 is proved in a slightly stronger form.

LEMMA 5.4. *Under Assumption* 3 *we have* $w_{A,B}(t) > 0$ *for all* $t \in [b_1, b_n] \setminus \{b_1, ..., b_n\}$ *and*

$$\operatorname{supp}(\mu_{A,B}) = [b_1, b_n] = [\tilde{b}_1, \tilde{b}_n]. \tag{5.25}$$

Proof. The lemma is an immediate consequence of the strict inequality in (5.4) in Proposition 5.1. □

5.4. The general case

In this subsection we show that Assumption 3, which has played a central role in §5.3, is actually superfluous for the proof of positivity of the measure $\mu_{A,B}$ in Theorems 1.6. For this purpose we have to revisit some definitions and results from §3.1 and §3.2.

If the polynomial $g(\lambda, t)$ in (3.1) is not irreducible, then it can be factorized into $m>1$ irreducible factors $g_{(l)}(\lambda, t)$, $l=1, ..., m$, of degree n_l as already stated in (3.1). For the partial degrees n_l we have $n_1+...+n_m=n$. Each polynomial $g_{(l)}(\lambda, t)$, $l=1, ..., m$, can be normalized in accordance with (3.4).

The m polynomial equations (3.2) define m algebraic functions $\lambda_{(l)}$, $l=1, ..., m$, and each of them has a Riemann surface $\mathcal{R}_{\lambda,l}$, $l=1, ..., m$, with n_l sheets over $\overline{\mathbb{C}}$ as its natural domain of definition. The solution λ of equation (3.1) consists of these m algebraic functions, and its domain of definition is the union (3.27) of the m Riemann surfaces $\mathcal{R}_{\lambda,l}$, $l=1,...,m$.

Each algebraic function $\lambda_{(l)}$, $l=1,...,m$, possesses n_l branches $\lambda_{l,k}$, $k=1,...,n_l$, which are assumed to be chosen analogously to Definition 3.7, but with a new form of indices. After (3.4) we have denoted by

$$j:\{(l,k):k=1,...,n_l,l=1,...,m\}\longrightarrow\{1,...,n\}$$

a bijection that establishes a one-to-one correspondence between the two types of indices that are relevant here. We may assume that this correspondence has been chosen in such a way that

$$b_{j(l,1)}\leqslant...\leqslant b_{j(l,n_l)}\quad\text{for each } l=1,...,m, \tag{5.26}$$

and in the new system of indices (3.6) in Lemma 3.3 takes the form

$$\lambda_{j(l,k)}(t)=\lambda_{l,k}(t)=a_{j(l,k),j(l,k)}-b_{j(l,k)}t+O\left(\frac{1}{t}\right)\quad\text{as } t\to\infty \tag{5.27}$$

for $k=1,...,n_l$, $l=1,...,m$.

We define

$$w_{A,B,l}(t):=\sum_{\substack{k=1\\ b_{j(l,k)}<t}}^{n_l}\frac{1}{2\pi i}\oint_{C_{l,k}}e^{\lambda_{l,k}(\zeta)+t\zeta}\,d\zeta\quad\text{for } l=1,...,m \tag{5.28}$$

with $C_{l,k}=C_{j(l,k)}$. From (5.28) it follows that in (1.13) and (1.14) in Theorem 1.6 we have

$$w_{A,B}(t)=\sum_{l=1}^{m}w_{A,B,l}(t). \tag{5.29}$$

Under Assumption 3 the new definitions remain consistent in a trivial way with $m=1$.

In the general proof of positivity of the measure $\mu_{A,B}$ the next proposition will take the role of Proposition 5.1.

PROPOSITION 5.5. (i) *For each* $l\in\{1,...,m\}$ *with* $n_l=1$ *we have*

$$w_{A,B,l}(t)=0\quad\textit{for all } t\in\mathbb{R}_+. \tag{5.30}$$

(ii) *For each* $l\in\{1,...,m\}$ *with* $n_l>1$ *we have*

$$w_{A,B,l}(t)\begin{cases}>0 & \textit{for all } t\in[b_{j(l,1)},b_{j(l,n_l)}]\setminus\{b_{j(l,1)},...,b_{j(l,n_l)}\},\\ =0 & \textit{for all } t\in\mathbb{R}_+\setminus[b_{j(l,1)},b_{j(l,n_l)}].\end{cases} \tag{5.31}$$

Each function $w_{A,B,l}$, $l=1,...,m$, *is the restriction of an entire function in each interval of* $[b_{j(l,1)},b_{j(l,n_l)}]\setminus\{b_{j(l,1)},...,b_{j(l,n_l)}\}$.

Proof. Equality (5.30) and the equality in the second line of (5.31) follow from (5.28) and the analogue of Lemma 4.1, which also holds for each complete set of branches $\lambda_{l,k}$, $k=1, ..., n_l$, of the algebraic function $\lambda_{(l)}$, $l=1, ..., m$. In the case of the second line in (5.31) we have also to take into consideration the ordering (5.26).

For the proof of the inequality in the first line of (5.31) we have to redo the analysis in the proofs of Lemmas 5.2 and 5.3, and of Proposition 5.1, but now with the role of the algebraic function λ, the Riemann surface $\mathcal{R}_\lambda$ and the branches λ_j, $j=1, ..., n$, taken over by $\lambda_{(l)}$, $\mathcal{R}_{\lambda,l}$ and $\lambda_{l,k}$, $k=1, ..., n_l$, respectively, for each $l=1, ..., m$ with $n_l>1$. It is not difficult to see that this transition is a one-to-one copying of all steps of the earlier analysis, and we will not go into further details. The inequality in the first line of (5.31) follows then together with (5.28) as an analogue of (5.4) in Proposition 5.1.

It follows from (5.28) that each $w_{A,B,l}$ is the restriction of an entire function in each interval in $[b_{j(l,1)}, b_{j(l,n_l)}]\setminus\{b_{j(l,1)}, ..., b_{j(l,n_l)}\}$ for $l=1, ..., m$. □

5.5. General proof of positivity

With (5.28) and Proposition 5.5 we are prepared for the proof of positivity without Assumption 3.

General proof of positivity. Since the discrete part (5.24) of the measure $\mu_{A,B}$ is positive, it remains only to show that the density function $w_{A,B}$ in (1.13) of Theorem 1.6 is non-negative in $[\tilde{b}_1, \tilde{b}_n]\setminus\{\tilde{b}_1, ..., \tilde{b}_n\}=[b_1, b_n]\setminus\{b_1, ..., b_n\}$. But this follows immediately from (5.31) and (5.30) in Proposition 5.5 together with (5.28). Notice that, because of Assumption 1 in §2, we have $\tilde{b}_j=b_j$ for $j=1, ..., n$. □

6. Summing up the proofs of Theorems 1.3 and 1.6

All assertions of Theorem 1.6, except for the positivity of the measure $\mu_{A,B}$, have been proved in §4, and after the proof of positivity in the last section, the proof of Theorem 1.6 is complete.

Theorem 1.3 is an immediate consequence of Theorem 1.6.

7. Proof of Proposition 1.2

The proof of Proposition 1.2 is given in two steps. In the first one, formulae (1.5) and (1.6) are verified. After that, in §7.2, it is shown that the density function $w_{A,B}(x)$ in (1.6) is positive for $b_1<x<b_2$. In the last subsection, representation (1.6) of the density function $w_{A,B}$ in Proposition 1.2 is compared with the corresponding result in [17].

7.1. Proofs of representations (1.5) and (1.6)

Representation (1.5) of the general structure of the measure $\mu_{A,B}$ follows as a special case from the analogous result (1.12) in Theorem 1.6. From (1.13) we further deduce that the density function $w_{A,B}$ in (1.5) can be represented as

$$w_{A,B}(x) = \frac{1}{2\pi i} \oint_{C_1} e^{\lambda_1(\zeta)+x\zeta}\, d\zeta \quad \text{for } b_1 < x < b_2, \tag{7.1}$$

with λ_1 being the branch of the algebraic function λ of degree 2 defined by the polynomial equation

$$g(\lambda,t) = \det(\lambda I-(A-tB)) = (\lambda+b_1t-a_{11})(\lambda+b_2t-a_{22})-|a_{12}|^2 = 0 \tag{7.2}$$

that satisfies

$$\lambda_1(t) = a_{11}-b_1t+O\left(\frac{1}{t}\right) \quad \text{as } t\to\infty. \tag{7.3}$$

Further, the integration path C_1 in (7.1) is a positively oriented Jordan curve that contains all branch points of the function λ in its interior. From (7.2) and (7.3) it follows that λ_1 is explicitly given by

$$\lambda_1(t) = \tfrac{1}{2}\left[(a_{22}+a_{11})-(b_2+b_1)t+\sqrt{[(a_{11}-a_{22})+(b_2-b_1)t]^2+4|a_{12}|^2}\,\right], \tag{7.4}$$

with the sign of the square root in (7.4) chosen in such a way that $\sqrt{\dots}\approx(b_2-b_1)t$ for t near ∞. Evidently, λ_1 has the two branch points

$$t_{1,2} = \frac{a_{22}-a_{11}}{b_2-b_1} \pm i\frac{2|a_{12}|}{b_2-b_1}. \tag{7.5}$$

The main task is now to transform the right-hand side of (7.1) into the more explicit expression in (1.6). In order to simplify the exponent in (7.1), we introduce a new variable v by the substitution

$$t(v) := \frac{a_{22}-a_{11}}{b_2-b_1} + \frac{2}{b_2-b_1}v, \quad v\in\mathbb{C}, \tag{7.6}$$

which leads to

$$\begin{aligned}(\lambda_1\circ t)(v)+xt(v) &= \frac{a_{11}(b_2-x)+a_{22}(x-b_1)}{b_2-b_1} + \frac{2x-(b_2+b_1)}{b_2-b_1}v+\sqrt{|a_{12}|^2+v^2} \\ &= \frac{a_{11}(b_2-x)+a_{22}(x-b_1)}{b_2-b_1} + g(v),\end{aligned} \tag{7.7}$$

with

$$g(v) := \frac{2x-(b_2+b_1)}{b_2-b_1}v+\sqrt{|a_{12}|^2+v^2}. \tag{7.8}$$

Notice that if x moves between b_1 and b_2, then the first term in the second line of (7.7) moves between a_{11} and a_{22}, and the coefficient in front of v in the second term moves between -1 and 1. The assumption made after (7.4) with respect to the square root transforms into $\sqrt{|a_{12}|^2+v^2}\approx v$ for v near ∞. It is evident that g is analytic and single-valued throughout $\overline{\mathbb{C}}\setminus[-i|a_{12}|,i|a_{12}|]$. From (7.7) and (7.1) we deduce the representation

$$w_{A,B}(x)=\frac{2}{b_2-b_1}\exp\biggl(\frac{a_{11}(b_2-x)+a_{22}(x-b_1)}{b_2-b_1}\biggr)\frac{1}{2\pi i}\oint_{C_1}e^{g(v)}\,dv, \tag{7.9}$$

where again C_1 is a positively oriented Jordan curve, which is contained in the ring domain $\mathbb{C}\setminus[-i|a_{12}|,i|a_{12}|]$. Shrinking this curve to the interval $[-i|a_{12}|,i|a_{12}|]$ yields that

$$\begin{aligned} w_{A,B}(x)=\frac{1}{(b_2-b_1)\pi}&\exp\biggl(\frac{a_{11}(b_2-x)+a_{22}(x-b_1)}{b_2-b_1}\biggr)\\ &\times\int_{-|a_{12}|}^{|a_{12}|}\exp\biggl(-i\frac{b_2+b_1-2x}{b_2-b_1}v\biggr)\bigl(e^{\sqrt{|a_{12}|^2-v^2}}-e^{-\sqrt{|a_{12}|^2-v^2}}\bigr)\,dv,\end{aligned} \tag{7.10}$$

and further that

$$\begin{aligned} w_{A,B}(x)=\frac{4}{(b_2-b_1)\pi}&\exp\biggl(\frac{a_{11}(b_2-x)+a_{22}(x-b_1)}{b_2-b_1}\biggr)\\ &\times\int_0^{|a_{12}|}\cos\biggl(\frac{b_2+b_1-2x}{b_2-b_1}v\biggr)\sinh\bigl(\sqrt{|a_{12}|^2-v^2}\bigr)\,dv,\end{aligned} \tag{7.11}$$

which proves formula (1.6).

7.2. The positivity of $w_{A,B}$

Since Proposition 1.2 is a special case of Theorem 1.6, and since the matrices A and B have been given in the special form of Assumption 3 in §5.1, the positivity of $w_{A,B}(x)$ for $b_1<x<b_2$ has in principle already been proved by Proposition 5.1. However, the prominence of the positivity problem in the BMV conjecture may justify an ad hoc proof for the special case of dimension $n=2$, which is simpler than the general approach in §5, and may also serve as an illustration for the basic ideas in this approach.

From (7.1) and (7.7)–(7.9), it follows that we only have to prove that

$$I_0:=\frac{1}{2\pi i}\oint_{C_1}e^{g(\zeta)}\,d\zeta=\frac{2}{\pi}\int_0^a\cos(bv)\sinh\bigl(\sqrt{a^2-v^2}\bigr)\,dv>0 \tag{7.12}$$

with the function g defined in (7.8), a and b being abbreviations for

$$a:=|a_{12}|\quad\text{and}\quad b:=b(x)=\frac{2x-(b_2+b_1)}{b_2-b_1}, \tag{7.13}$$

respectively, and C_1 being a positively oriented integration path in the ring domain $\mathbb{C}\setminus[-ia,ia]$.

Obviously, we have $-1<b(x)<1$ for $b_1<x<b_2$. The value I_0 of the second integral in (7.12) depends evenly on the parameter b, and I_0 is obviously positive for $b=0$. Consequently, we can, without loss of generality, restrict our investigation to values of $x\in(b_1,b_2)$ that correspond to values $b\in(-1,0)$, and they are $b_1<x<\frac{1}{2}(b_1+b_2)$.

For a fixed value $x\in\left(b_1,\frac{1}{2}(b_1+b_2)\right)$ we now study the behavior of the function g of (7.8) in $\mathbb{C}\setminus[-ia,ia]$. Because of the convention with respect to the sign of the square root in (7.8), we have

$$g(z)\approx(1+b)z \quad \text{for } z\approx\infty. \tag{7.14}$$

The function $\operatorname{Im} g$ is continuous in $\mathbb{C}$, harmonic in $\mathbb{C}\setminus[-ia,ia]$, we have

$$\operatorname{Im} g(\bar{z})=-\operatorname{Im} g(z) \quad \text{for } z\in\mathbb{C},$$

and

$$\operatorname{Im} g(z)=b\operatorname{Im} z\begin{cases} <0 & \text{for } z\in(0,ia],\\ >0 & \text{for } z\in[-ia,0).\end{cases} \tag{7.15}$$

From (7.14), (7.15), $1+b>0$ and the harmonicity of $\operatorname{Im} g$, we deduce that the set

$$\{z:\operatorname{Im} g(z)=0\}=\mathbb{R}\cup\gamma \tag{7.16}$$

implicitly defines an analytic Jordan curve γ, which is contained in $\mathbb{C}\setminus[-ia,ia]$. We parameterize this curve by $\gamma:[0,2\pi]\to\mathbb{C}$ in such a way that it is positively oriented in $\mathbb{C}$ and that

$$\gamma|_{(0,\pi)}\subset\{z:\operatorname{Im} z>0\}, \quad \gamma(0)=:r_0>0 \quad \text{and} \quad \gamma(2\pi-t)=\overline{\gamma(t)} \text{ for } t\in[0,\pi]. \tag{7.17}$$

From (7.16) it follows that g is real on γ. Further, we have

$$(g\circ\gamma)'(t)<0 \quad \text{for } t\in(0,\pi). \tag{7.18}$$

Indeed, if we set $D_+:=\operatorname{Ext}(\gamma)\cap\{z:\operatorname{Im} z>0\}$ and $D_-:=\operatorname{Int}(\gamma)\cap\{z:\operatorname{Im} z>0\}$, then it follows from (7.14), $1+b>0$, (7.15) and (7.16) that

$$\operatorname{Im} g(z)\begin{cases} >0 & \text{for } z\in D_+,\\ <0 & \text{for } z\in D_-,\end{cases}$$

and with the harmonicity of $\operatorname{Im} g$ we deduce that

$$\left(\frac{\partial}{\partial n}\operatorname{Im} g\right)\circ\gamma(t)<0 \quad \text{for } t\in(0,\pi),$$

where $\partial/\partial n$ denotes the normal derivative on γ pointing into D_-. The inequality in (7.18) then follows by the Cauchy–Riemann differential equations and the fact that $g\circ\gamma = \operatorname{Re} g\circ\gamma$.

With the Jordan curve γ and the inequality in (7.18) we are prepared to prove the positivity of the integral I_0 in (7.12). Using γ as integration path in the first integral in (7.12) yields that

$$\begin{aligned} I_0 &= \frac{1}{2\pi i}\int_0^{2\pi} e^{g\circ\gamma(t)}\gamma'(t)\,dt \\ &= \frac{1}{\pi}\operatorname{Im}\int_0^{\pi} e^{g\circ\gamma(t)}\gamma'(t)\,dt \\ &= \frac{1}{\pi}\operatorname{Im}[e^{g\circ\gamma(t)}\gamma(t)]_0^{\pi} - \frac{1}{\pi}\operatorname{Im}\int_0^{\pi}(g\circ\gamma)'(t)e^{g\circ\gamma(t)}\gamma(t)\,dt \\ &= -\frac{1}{\pi}\int_0^{\pi}(g\circ\gamma)'(t)e^{g\circ\gamma(t)}\operatorname{Im}\gamma(t)\,dt > 0. \end{aligned} \tag{7.19}$$

Indeed, the second equality in (7.19) is a consequence of the symmetry relations

$$(g\circ\gamma)(t) = (g\circ\gamma)(2\pi - t), \quad \gamma'(t) = -\overline{\gamma'(2\pi - t)}, \quad \text{and} \quad \gamma(t) = \overline{\gamma(2\pi - t)}$$

for $t\in[0,2\pi)$. The next equality follows from partial integration, and the last equality is a consequence of $\operatorname{Im}\gamma(0) = \operatorname{Im}\gamma(\pi) = 0$ and $\operatorname{Im}(g\circ\gamma)(t) = 0$ for $t\in[0,2\pi)$. Finally, the inequality in (7.19) is a consequence of (7.18) together with $\operatorname{Im}\gamma(t) > 0$ for $t\in(0,\pi)$.

With (7.19) we have verified that $w_{A,B}(x) > 0$ for all $x\in(b_1,b_2)$, which completes the proof of Proposition 1.2.

7.3. A comparison with the solution in [17]

In [17, Formulae (2.13)–(2.16)] an explicit representation for the measure $\mu_{A,B}$ has been proved for the case of dimension $n=2$, in which the expression of the density function $w_{A,B}$ differs considerably in its appearance from representation (1.6) in Proposition 1.2; it reads([2]) as

$$w_{A,B}(x) = \exp\left(\frac{a_{11}(b_2 - x) + a_{22}(x - b_1)}{b_2 - b_1}\right) G_{12}(x) \tag{7.20}$$

with

$$G_{12}(x) = \sum_{j=1}^{\infty} \frac{|a_{12}|^{2j}}{j!(j-1)!} \frac{(b_2 - x)^{n-1}(x - b_1)^{n-1}}{(b_2 - b_1)^{2n-1}}, \quad b_1 < x < b_2, \tag{7.21}$$

([2]) Formula (2.15) of [17], which is reproduced here as (7.21), contains a misprint; there is written erroneously $2n+1$ instead of $2n-1$ in the exponent of the denominator. The correction can easily be verified by following its derivation starting from [17, (2.11)].

where we use the terminology from Proposition 1.2. The representations (7.21) and (1.6) have not only a rather different appearance, they have also been obtained by very different approaches. However, they are identical, as will be shown in the next lines. We have to show that

$$G_{12}(x) = \frac{4}{(b_2-b_1)\pi}\int_0^{|a_{12}|}\cos\left(\frac{b_2+b_1-2x}{b_2-b_1}u\right)\sinh\left(\sqrt{|a_{12}|^2-u^2}\,\right)du \tag{7.22}$$

for $b_1<x<b_2$.

We use the same abbreviations a and b as in (7.13). From

$$\cos(bu)\sinh\left(\sqrt{a^2-u^2}\,\right) = \sum_{j=0}^{\infty}\sum_{k=1}^{\infty}\frac{(-1)^j b^{2j}}{(2j)!(2k-1)!}\frac{u^{2j}(a^2-u^2)^k}{\sqrt{a^2-u^2}}$$

and

$$\int_0^a \frac{u^{2j}(a^2-u^2)^k}{\sqrt{a^2-u^2}}\,du = a^{2(j+k)}\frac{\Gamma\left(j+\frac{1}{2}\right)\Gamma\left(k+\frac{1}{2}\right)}{(j+k)!} = \pi a^{2(j+k)}\frac{(2j)!(2k)!}{2^{2(j+k)}(j+k)!j!k!},$$

we deduce that

$$\begin{aligned}\int_0^a \cos(bu)\sinh\left(\sqrt{a^2-u^2}\,\right)du &= \pi\sum_{j=0}^{\infty}\sum_{k=1}^{\infty}(-1)^j b^{2j}a^{2(j+k)}\frac{4^{-(j+k)}}{(j+k)!j!(k-1)!}\\ &= \pi\sum_{n=1}^{\infty}\frac{a^{2n}}{4^n n!(n-1)!}\sum_{j=0}^{n-1}(-1)^j\frac{(n-1)!}{j!(n-j-1)!}b^{2j}\\ &= \frac{\pi}{4}\sum_{n=1}^{\infty}\frac{a^{2n}}{n!(n-1)!}\left(\frac{1-b^2}{4}\right)^{n-1}\\ &= \frac{\pi}{4}\sum_{n=1}^{\infty}\frac{|a_{12}|^{2n}}{n!(n-1)!}\frac{(b_2-x)^{n-1}(x-b_1)^{n-1}}{(b_2-b_1)^{2(n-1)}}.\end{aligned} \tag{7.23}$$

The last equality in (7.23) follows from

$$\frac{1-b^2}{4} = \frac{1}{4}\left(1-\left(\frac{b_2+b_1-2x}{b_2-b_1}\right)^2\right) = \frac{(b_2-x)(x-b_1)}{(b_2-b_1)^2}.$$

With (7.23), the identity (7.22) is proved.

Acknowledgements. The author is very grateful to Pierre Moussa for his enthusiastic welcome of the first version of the present paper and for valuable hints to interesting earlier publications, to Peter Landweber for many improvements in the manuscript, to the referee for further improvements, and last but not least, to Elliott H. Lieb and Robert Seiringer for a valuable discussion, and especially for the suggestion to simplify and shorten the proof of the conjecture dramatically by going for a direct verification of formulae (1.12)–(1.14) in Theorem 1.6, as is now done in §4. In the original version of the paper [20] these formulae were proved by a lengthy, asymptotic analysis of the function (1.1) with a subsequent use of the Post–Widder inversion formulae for Laplace transforms.

References

[1] Bessis, D., Moussa, P. & Villani, M., Monotonic converging variational approximations to the functional integrals in quantum statistical mechanics. *J. Math. Phys.*, 16:11 (1975), 2318–2325.

[2] Burgdorf, S., Sums of Hermitian squares as an approach to the BMV conjecture. *Linear Multilinear Algebra*, 59 (2011), 1–9.

[3] Collins, B., Dykema, K. J. & Torres-Ayala, F., Sum-of-squares results for polynomials related to the Bessis–Moussa–Villani conjecture. *J. Stat. Phys.*, 139 (2010), 779–799.

[4] Donoghue, W. F., Jr., *Monotone Matrix Functions and Analytic Continuation*. Die Grundlehren der mathematischen Wissenschaften, 207. Springer, New York, 1974.

[5] Farkas, H. M. & Kra, I., *Riemann Surfaces*. Graduate Texts in Mathematics, 71. Springer, New York, 1992.

[6] Fleischhack, C. & Friedland, S., Asymptotic positivity of Hurwitz product traces: two proofs. *Linear Algebra Appl.*, 432:6 (2010), 1363–1383.

[7] Grafendorfer, G., *Hardening of the BMV Conjecture*. Diploma Thesis, Technische Universität Wien, Wien, 2007. http://www.math.ethz.ch/~ggeorg/files/thesis.pdf.

[8] Hägele, D., Proof of the cases $p\leqslant 7$ of the Lieb–Seiringer formulation of the Bessis–Moussa–Villani conjecture. *J. Stat. Phys.*, 127 (2007), 1167–1171.

[9] Hansen, F., Trace functions as Laplace transforms. *J. Math. Phys.*, 47 (2006), 043504, 11 pp.

[10] Hillar, C. J., Advances on the Bessis–Moussa–Villani trace conjecture. *Linear Algebra Appl.*, 426 (2007), 130–142.

[11] Hillar, C. J. & Johnson, C. R., On the positivity of the coefficients of a certain polynomial defined by two positive definite matrices. *J. Stat. Phys.*, 118 (2005), 781–789.

[12] Johnson, C. R., Leichenauer, S., McNamara, P. & Costas, R., Principal minor sums of $(A+tB)^m$. *Linear Algebra Appl.*, 411 (2005), 386–389.

[13] Klep, I. & Schweighofer, M., Sums of Hermitian squares and the BMV conjecture. *J. Stat. Phys.*, 133 (2008), 739–760.

[14] Landweber, P. S. & Speer, E. R., On D. Hägele's approach to the Bessis–Moussa–Villani conjecture. *Linear Algebra Appl.*, 431:8 (2009), 1317–1324.

[15] Lieb, E. H. & Seiringer, R., Equivalent forms of the Bessis–Moussa–Villani conjecture. *J. Stat. Phys.*, 115 (2004), 185–190.

[16] — Further implications of the Bessis–Moussa–Villani conjecture. *J. Stat. Phys.*, 149 (2012), 86–91.

[17] Mehta, M. L. & Kumar, K., On an integral representation of the function $\mathrm{Tr}[\exp(A-\lambda B)]$. *J. Phys. A*, 9 (1976), 197–206.

[18] Moussa, P., On the representation of $\mathrm{Tr}(e^{(A-\lambda B)})$ as a Laplace transform. *Rev. Math. Phys.*, 12 (2000), 621–655.

[19] Reed, M. & Simon, B., *Methods of Modern Mathematical Physics*. IV. *Analysis of Operators*. Academic Press, New York, 1978.

[20] Stahl, H. R., Proof of the BMV conjecture. Preprint, 2011. arXiv:1107.4875 [math.CV].

[21] Widder, D. V., *The Laplace Transform*. Princeton Mathematical Series, 6. Princeton University Press, Princeton, NJ, 1941.

Herbert R. Stahl
was at
Beuth Hochschule/FB II
Luxemburger Str. 10
DE-13 353 Berlin
Germany

Received August 17, 2012